Essays on Attention and Information in Digital Media Markets

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2026

vorgelegt von

M.Sc. Lara Marie Berger

aus Lich

Referent:	Prof. Dr. Johannes Münster
Korreferent:	Prof. Dr. Bettina Rockenbach
Tag der Promotion:	06.07.2026

Acknowledgements

I am deeply grateful to my primary advisor, Johannes Münster, for his continuous support, guidance, and encouragement throughout this thesis. Your thoughtful feedback, countless discussions, and trust in my ideas have helped this dissertation grow and have made me a much better researcher. I have learned so much from you.

I am also deeply grateful to my second advisor, Bettina Rockenbach. You have always supported me and encouraged me to see the bigger picture and to tackle the big questions. Thank you.

I am deeply indebted to my wonderful co-authors of chapters 3 and 4, Anna Kerkhof, Felix Mindl, Johannes Münster and Nikola Noske. Working with you has made writing this thesis a truly enjoyable experience, and I am grateful for the many ways in which our collaboration has strengthened these papers.

I am also very grateful for helpful discussions and suggestions from Maja Adena, Raphaela Andres, Kai Barron, Irene Bertschek, Michael Brottrager, Felix Chopra, Gönül Dogan, Oliver Falck, Marcel Garz, Susanna Grundmann, Julian Harke, Sören Harrs, Tiffany Hopkins, Manfred Küllinger, Yero Ndiaye, Arianna Ornaghi, Andrea Prat, Dominik Rehse, Lukas Reinhardt, Fabian Roeben, Christopher Roth, Paula Scholz, Frederik Schwerter, Marta Serra-Garcia, Louis Strang, and participants at various seminars and conferences.

Johanna Gans, Annecke Sievers, Anne-Sophie Hille, Olga Andreeva, Lara Mai, and Aleksander Vasic provided excellent research assistance. Thank you so much for your help.

I also thank the Center for Social and Economic Behavior (C-SEB) at the University of Cologne, the ECONtribute Cluster of the Universities of Bonn and Cologne, and the Joachim Herz Stiftung for their financial support.

Finally, I thank my entire family and best friends, especially my husband Georg as well as Judith, Julian, Louisa, and Nora. You believed in me, supported me in the most difficult moments, and helped me refocus on what really matters. Thank you.

I dedicate this thesis to the greatest joy in my life, my son Frederik. You have brought indescribable love and perspective into my life while I was writing this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"A wealth of information creates a poverty of attention."

Herbert Simon (1971)

Digital media have made information access easier and faster than ever before. But does this mean that the public is better informed? The answer is not obvious. More than half a century ago, Simon (1971) warned that in an information-rich world the binding constraint is not the supply of information but the allocation of attention. On one hand, digital technologies, like algorithms and large language models, can help individuals find, summarize, and interpret information. On the other hand, they also vastly increase the amount of content competing for limited attention (Wu, 2016). Distortions in how the public is informed are not new (e.g., Gentzkow and J. M. Shapiro, 2010; Mullainathan and Shleifer, 2005), but the digital transformation of media markets has intensified them in ways that make this question particularly pressing today. When information is abundant, but attention remains scarce, which information reaches people, how is it framed, how is it interpreted, and how is it acted upon?

This dissertation shows that the answers to this question run largely through the digital media markets that now mediate the path from information to audience. These markets shape what is produced, how it travels, and how it is processed. How accurately information travels from its source to the public is not a narrow concern. What people come to know and believe ultimately shapes their emotions and their economic decisions. The overarching research question of the thesis is therefore how the digitization of media markets shapes the information that reaches the public, and how the production and processing of that information can be supported so that abundance translates into knowledge rather than distortion. The three chapters approach this question in two steps. Chapter 2 is mainly diagnostic. It traces how the digitization of news markets reshapes central dimensions of information quality, the sentiment and accuracy of headlines, following it from what journalists produce to what readers take away. Chapters 3 and 4 then turn constructive, testing concrete interventions that can partly counteract the problems this environment creates. On the demand side this involves equipping audiences to evaluate content better (Chapter 3) and on the supply side equipping journalists to communicate it more accurately (Chapter 4). A central theme running across all three is the tension between attention and reliable inference, as the same features that make information engaging and persuasive may also distort learning, beliefs, and ultimately decisions.

Chapter 2, *"How digital media markets amplify news sentiment"*, examines how digital attention incentives reshape the emotional tone of news, and traces their effects across both sides of the market, from what journalists produce to how readers respond to it. If attention is scarce, information does not reach audiences simply because it is available, relevant, or accurate. It must first be noticed. This creates incentives for news producers to select and present information in ways that stand out in competitive digital environments. An extensive descriptive literature has documented negativity biases in news (e.g., Soroka et al., 2019; Trussler and Soroka, 2014). This chapter complements it with causal evidence on their market-based determinants. The chapter studies this mechanism in the context of news headlines, asking whether digital attention incentives change the emotional tone of journalistic content and whether these changes affect how readers engage with and learn from news.

To answer this question, the chapter combines large-scale text analysis with two linked randomized experiments involving professional journalists and readers. The descriptive analysis documents that online headlines are substantially more emotional and more negative than comparable print headlines. The journalist experiments then isolate one mechanism behind this pattern: when journalists are rewarded for audience engagement, they choose and write more emotional and more negative headlines. The linked reader experiment shows why these incentives matter. Emotional and competitive headline environments increase engagement, but reduce factual learning. Chapter 2 thereby highlights a market-based channel through which digital attention incentives reshape what information audiences encounter. It also underlines a central trade-off in digital media markets: the same features that make information more likely to attract attention may also make it less effective at transmitting factual knowledge.

Having documented how attention incentives reshape news, a natural next question is what can be done about it. **Chapter 3**, *"Debunking 'fake news' on social media: Immediate and short-term effects of fact-checking and media literacy interventions"* (joint with Anna Kerkhof, Felix Mindl, and Johannes Münster), begins on the demand side of digital information environments. Once information reaches individuals, it still has to be interpreted and evaluated. This is not a trivial task in environments where accurate and misleading content compete for attention and often appear in similar formats (e.g., Allcott and Gentzkow, 2017; D. M. Lazer et al., 2018). Fact-checking and media literacy have each been studied as remedies for belief in misinformation, but rarely within the same design. This chapter provides the first clean comparison of the two and contributes to a still thin body of evidence on media literacy in particular. The chapter therefore asks whether and which interventions can help individuals assess the credibility of information more accurately, and whether such interventions improve belief formation only for specific corrected claims or also more generally.

To answer this question, we expose a sample of German residents to real false and correct posts circulating on Facebook about Covid-19 vaccines and nutrition in a large pre-registered survey experiment. In this framework, we cleanly compare two prominent remedies discussed in policy debates: explicit fact-checking of selected posts, and a brief media-literacy intervention that teaches users how to evaluate content for themselves (building on Barrera et al., 2020; A. M. Guess et al., 2020). We measure beliefs, factual knowledge, and attitudes immediately after the intervention and again about two weeks later. We find that fact-checking corrects mainly the specific fakes that it addresses. The media-literacy intervention, by contrast, helps participants distinguish between fakes and facts more generally, both right away and two weeks on, and increases their critical engagement with the posts. In an environment where only a small share of misleading content can ever be fact-checked, equipping consumers with general skills to navigate that environment thereby emerges as the more scalable response. Chapter 3 thus shows that the processing of information, not only its supply, can be improved, and that relatively low-cost interventions can help individuals navigate digital environments.

Chapter 3 showed that audiences can be equipped to evaluate information more accurately. The same question on the supply side is whether producers, too, can be equipped to communicate it more accurately in the first place. **Chapter 4**, *"Improving science literacy in the newsroom: Experimental evidence"* (joint with Anna Kerkhof and Nikola Noske), takes up this question. The chapter focuses on a domain where the tension between attention and accuracy is particularly visible: the reporting of scientific findings (Dahlstrom, 2021). Scientific studies are typically technical, qualified, and probabilistic, while news markets reward content that is simple, vivid, and easy to grasp. Under tight deadlines, economic pressure, and often limited training in statistical methods, even well-intentioned journalists can produce reports that distort the underlying evidence (Oxman et al., 2022; Sumner et al., 2016). While the frequent inaccuracy of science reporting is well documented, little rigorous work tests what can be done about it. This chapter provides causal evidence on a supply-side remedy, and is among the few experiments conducted with professional journalists rather than with laypeople. The chapter therefore asks whether a brief, low-cost intervention can equip journalists themselves to communicate scientific findings more accurately, and whether it also changes the way they evaluate science-related reporting more generally.

To test this, we developed a seven-minute educational video (in collaboration with a journalism school) that encourages journalists to think critically about core principles of statistical interpretation and science communication. In a pre-registered online experiment with 260 professional journalists, half of the participants are randomly assigned to watch this video before being asked to write a headline for a real scientific study and to evaluate a published news article about that study. Treated journalists are 28 percentage points more likely to write a factually accurate headline than their counterparts in the control group, and they more frequently identify specific error types such as statistical misinterpretations and confusions between correlation and causation. Exploratory evidence from tracking a subset of participants' published work in the months following the experiment further suggests that the effect may extend to their real-world reporting. Chapter 4 thus complements Chapter 3 by showing that producers, too, can be supported in ways that improve the quality of information at its source. Taken together, the two chapters illustrate that the trade-off between attention and reliable inference documented in Chapter 2 is not an unavoidable feature of digital media markets, but can be partly counteracted by scalable interventions on both the demand and the supply side of the market.

The three chapters together show that digital information environments do not merely affect how much information is available. They shape the path along which information travels before it can inform the public: how it is produced, which messages attract attention, and how they are interpreted and acted upon. The dissertation therefore contributes to understanding when information abundance translates into better public knowledge, and when it instead creates distortions in learning. It also shows that these distortions are not inevitable, but can be partly addressed through interventions that improve information processing. Information abundance, this dissertation therefore suggests, is neither an unambiguous benefit nor an inevitable cost. Its consequences for public understanding depend on the design of the markets, technologies, and interventions that allocate the scarce resource Simon (1971) long ago identified as the binding one: attention.

Contribution to the co-authored chapters

Following the graduation regulations of the Faculty of Management, Economics, and Social Sciences at University of Cologne, this section explicates how I contributed to the co-authored chapters of this thesis.

The chapter "*Debunking 'fake news' on social media: Immediate and short-term effects of fact-checking and media literacy interventions*" (**Chapter 3**) is joint work with Anna Kerkhof, Felix Mindl, and Johannes Münster. All authors contributed equally to the development of the research idea and the design of the experiments. Felix and I implemented the experiments and conducted most of the data analysis. Anna prepared most of the initial draft, while Felix and I were primarily responsible for preparing the revised and published version.

Chapter 4, "*Improving science literacy in the newsroom: Experimental evidence*", is joint work with Anna Kerkhof and Nikola Noske. All authors contributed to the development of the research idea and the experimental design. I was responsible for the implementation of the experiment and the data collection. Nikola and I jointly conducted the data analysis and prepared both the initial and the revised draft of the paper.

Chapter 2

How digital media markets amplify news sentiment

Paper Information

Authors: Lara Marie Berger

Status: Working paper.

Abstract

This paper investigates how digital attention incentives influence the emotional tone of news headlines and affect the transmission of information to audiences. An analysis of more than 330,000 newspaper articles demonstrates that online headlines are markedly more emotional and negative than their print counterparts. Experiments with professional journalists reveal that stronger attention incentives, in the form of both performance-based pay and headline-level competition, causally lead to more emotional and negative headline choices. Connected experiments with readers indicate that such attention-driven environments increase engagement, but reduce factual learning. The paper thereby identifies a market-based mechanism of media bias, wherein commercial pressure to capture attention systematically reshapes how information is framed and what audiences take away from it.

2.1 Introduction

"If it bleeds, it leads." This old newsroom maxim captures a longstanding regularity: emotionally charged, and especially negative, stories disproportionately attract audience attention[1]. How news is framed shapes what citizens know about the economy (Snyder and Strömberg, 2010), how they form expectations (Carroll, 2003; Larsen et al., 2021), and how they evaluate the institutions that govern them (DellaVigna and Kaplan, 2007; Enikolopov et al., 2011). The economic conditions under which news is now produced have, however, changed profoundly in the past decades. Whereas print newspapers sold bundles of articles in which only a few front-page headlines competed for attention, digital and social media expose every headline to direct competition in vast online feeds. Audience engagement is measured in real time and increasingly determines which stories get amplified, how journalists are evaluated and promoted, and, in some cases, how they are paid.[2] At the same time, audiences themselves respond more strongly to emotionally salient and especially negative information (Gambetti et al., 2023; Rozin and Royzman, 2001; Soroka et al., 2019), which sharpens the returns to producing such content. Recent advances in generative AI may intensify these dynamics further by sharply reducing the cost of producing competing content. If commercial pressure to capture attention systematically shifts how information is framed, the consequences could extend well beyond newsrooms. It may distort economic perceptions, lower the quality of public deliberation, and make the media environment less useful as a basis for informed democratic choice.

This paper asks whether digital attention incentives systematically alter the information that journalists produce and the information that readers absorb. I focus in particular on two features that have become defining of digital news markets and that I jointly refer to as *attention incentives*: performance-based pay tied to reader engagement (*monetary attention incentives*), and direct headline-level competition (*competitive attention incentives*). Answering this question requires evidence on three connected pieces: (i) whether the digital shift is associated with a real change in how news is framed, (ii) whether that change is causally driven by attention incentives, and (iii) whether the resulting framing meaningfully affects what audiences learn. I therefore combine large-scale text analysis of newspaper headlines with two linked

[1]See, for example, evidence on negativity and positivity biases in news coverage (Bleich and Veen, 2021; Farnsworth and Lichter, 2011; Garz, 2014; Goidel and Langley, 1995; Hofstetter and Dozier, 1986; Kayser and Peress, 2021; Ryu, 1982; Soroka, 2016; Soroka et al., 2018; Soroka and Krupnikov, 2021).

[2]Many outlets monitor real-time engagement metrics and increasingly use them in editorial decisions; some have also experimented with linking remuneration to article performance (see e.g. Infosperber and The Guardian).

randomized experiments involving professional journalists and a representative reader sample.

The first step is to establish that there is a real-world pattern to be explained. If online and offline headlines did not in fact differ in tone, there would be little to attribute to attention incentives in the first place. I therefore fine-tune a machine-learning sentiment classifier on more than 330,000 economic news headlines from major German newspapers and compare the online and offline editions of the same outlets. Online headlines are on average 18 percentage points more likely to be emotional than offline headlines ($p < 0.01$), corresponding to 0.37 standard deviations relative to the offline group. Average tone is negative in both settings, but more strongly so online: on a sentiment scale from -1 to $+1$, online headlines are 0.10 points more negative than offline headlines (0.14 standard deviations, $p < 0.01$). The online-offline gap in emotionality is robust across alternative classification methods, human-coded subsamples, and non-economic news, and it also replicates in data from the *New York Times*.

Further, three additional patterns in the descriptive data are consistent with attention pressure as a candidate driver of this gap. First, within the print sample, front-page headlines, which face stronger competition for readers' attention than headlines printed inside the newspaper, are on average between 0.07 and 0.26 standard deviations more emotional and between 0.02 and 0.18 standard deviations more negative across all outlets ($p < 0.01$). Second, a 2014 legal dispute that temporarily removed *Die Welt* from several major German digital news aggregators is associated with a decline in online headline emotionality at that outlet (back-of-the-envelope difference-in-differences estimates of 0.08 to 0.13 standard deviations), consistent with reduced headline-level competition lowering the returns to emotional framing. Third, for two of the three outlets that introduce a digital paywall during the sample period, the emotionality of online headlines rises relative to the print edition following the introduction, consistent with paywalls strengthening the returns to capturing attention online.

The descriptive evidence thus documents a robust online-offline gap and points to attention pressure as a plausible driver, but it cannot identify a causal mechanism. Online and offline editions differ in many ways beyond attention pressure: audiences, topics, editorial workflows, and time horizons all vary across formats. To attribute a tonal gap specifically to journalists' incentives, I need variation in those incentives that is exogenous to the other features of digital markets. This is what the journalist experiment provides. In a pre-registered online experiment with $N = 402$ professional journalists, compensation is randomly varied to mirror the differences between attention incentives in *digital* media markets (treatment groups I and II) versus *analogue* media markets (control group). Each journalist produces a total of ten headlines for two randomly chosen real-world news articles. Journalists in the treatment groups are paid based on how many readers in the connected readers' experiment either click on their headlines or subscribe to a related newsletter after seeing them, while those in the control group receive a flat rate regardless of their headlines' performance.[3] In addition, the degree of competitive attention incentives at the headline level is varied within-subjects.[4]

When remuneration depends on audience engagement, journalists systematically select and write more emotional and more negative headlines for otherwise identical articles. In particular, journalists in the *PayPerClick* group are 12 percentage points more likely to produce an emotional headline ($p < 0.01$) than those in the control group, corresponding to 0.25 standard deviations in the control group. Journalists in the *PayPerAbo* group are 7 percentage points more likely to produce an emotional headline (0.14 standard deviations, $p < 0.01$). These differences arise entirely because journalists in the treatment groups produce more negative headlines than their counterparts in the control group. A supplemental experiment conducted earlier corroborates this result. Additionally, the within-subject manipulation shows that competitive attention incentives at the headline level independently intensify negative headline framing. Exploratory analyses of the subset of self-written headlines further reveal that attention incentives not only shift tone, but also affect quality more broadly. Both the pay-per-click and the pay-per-subscription scheme increase factual mistakes relative to the fixed-pay baseline. These results suggest that stronger attention incentives compromise not only emotional balance, but also accuracy in reporting.

A shift in headline tone is economically meaningful only if it changes what readers actually learn and do. Whether more emotional headlines hurt, help, or leave audiences indifferent cannot be inferred from supply-side patterns alone, and it determines whether the documented incentive mechanism is a welfare-relevant distortion or a stylistic change without consequence. I therefore conducted a linked, pre-registered reader experiment with a representative sample of Germans ($N =$ 1,617),[5] designed so that the headlines produced by the journalists in the experiment

[3] Although explicit pay-per-click contracts are rare in journalistic practice, attention metrics increasingly shape newsroom evaluations, promotion prospects, and assignment decisions. Freelance and platform-based creators ("news influencers") are often directly remunerated based on audience engagement. The experimental variation therefore captures a scalable and empirically relevant mechanism, even if the institutional form of incentives differs across settings.

[4] That is, journalists are informed whether the headlines they produce will be presented to readers either without a competing headline or alongside one competing headline.

[5] This sample is representative of the German population in terms of age, gender, and state of residence.

above are the stimuli shown to the readers here. This linkage ensures that the engagement effects estimated on the reader side correspond to the same headlines whose supply was experimentally manipulated. Readers are randomly exposed to headlines that vary in emotionality and headline-level competition, and are then observed as they decide which articles to click on, whether to subscribe to related newsletters, how much time to spend with the news, and what they subsequently report about the factual content of the articles, their emotions, and their agreement with polarizing statements.

Because only around 30 percent of displayed headlines in this experiment are actually clicked,[6] I distinguish between intention-to-treat (ITT) effects of assignment to more emotional and competitive headline environments and instrumental-variable (IV) estimates that capture the effect for readers who click because of that assignment.

The ITT estimates show that exposure to emotional headlines in competitive settings increases engagement (by raising the probability to click by 18 percentage points ($p < 0.01$) and time spent with news by about 8 seconds ($p < 0.01$)), but also reduces factual accuracy. Readers' answers in subsequent knowledge questions are on average 0.07 standard deviations further away from the truth ($p = 0.038$). The IV estimates indicate substantially larger effects among compliers. For readers induced to click by emotional and competitive headlines, time spent with news rises by about 73 seconds ($p < 0.01$), while factual error increases by 0.57 standard deviations ($p = 0.081$).[7] In substantive terms, this corresponds, for example, to misperceiving the current level of core inflation by 0.92 percentage points in the ITT specification and by 2.96 percentage points in the IV specification.

Taken together, the reader results document a clear trade-off in information transmission: attention-incentivized headline environments draw readers in (more clicks, more time spent) but distort what they take away (less accurate factual recall). In contrast, I do not find statistically detectable short-run effects on reported mood or on agreement with polarizing statements. Two features of the design likely explain these nulls. First, the experimental stimuli are deliberately mild: the "emotional" headlines are factually accurate, agency-style framings with slight negative or positive valence, not the extreme tabloid material that has motivated much of the prior literature on news exposure and affect. Second, exposure is short and one-off, whereas the affective consequences of negative news documented elsewhere typically arise under repeated, sustained exposure. The headline-level effects on factual accuracy I document can therefore be read as a conservative lower

[6] Participants are not monetarily incentivized to click on the headlines.

[7] All results reported here are comparisons of the group of readers with strong attention incentives (headline-level competition *and* emotionality) and low attention incentives (no competition and no emotionality).

bound on the welfare implications of attention-driven news markets. Cumulative exposure to systematically more emotional and more negative headlines in real-world digital environments is likely to produce stronger downstream effects, possibly also on readers' affect, trust in institutions, and the quality of citizens' economic and political beliefs.

Together, the results offer a coherent picture of how digital attention markets affect the production and consumption of news headlines. Descriptive evidence shows that online headlines are more emotional; experiments demonstrate that attention incentives induce journalists to such production patterns; and reader data reveal that this distorts information transmission to audiences. These findings highlight a market-based channel of media bias that is distinct from the widely discussed political slant or ownership influence. Even without ideological motives, commercial incentives to capture attention systematically reshape how information is framed and processed.[8] This emotional amplification can, for example, induce misinformed economic decisions, erode trust in the media, and shift public perceptions away from underlying realities.

This paper relates to three strands of research at the intersection of media economics, attention, and digitization. A first line of work documents systematic distortions in the emotional tone of media coverage, often referred to as negativity or positivity biases. Studies in this area identify persistent deviations in the valence of news relative to objective benchmarks[9] and show that audiences react more strongly to emotionally charged content than to neutral information (Leung and Strumpf, 2023; Robertson et al., 2023; Soroka et al., 2019; Trussler and Soroka, 2014; Watson et al., 2024). While this literature highlights the prevalence and consequences of emotional framing, most evidence is descriptive or correlates negative tone with higher demand (Arango-Kure et al., 2014; Dertwinkel-Kalt et al., 2022). What remains largely unresolved is the *causal* role of market incentives. This paper addresses that gap by experimentally varying the strength of attention incentives faced by professional journalists and by linking these supply-side responses to audience engagement and learning on the demand side.

The study also connects to recent experimental work on incentives in communication. In the closest antecedent, Serra-Garcia (2025) experimentally varies professional copywriters' monetary incentives to attract attention in science communication and

[8]This complements recent evidence by Angelucci and Prat (2024) who show that inequalities in news discernment, rather than partisan bias alone, drive differences in how accurately citizens identify true political news.

[9]See for instance Bleich and Veen (2021), Farnsworth and Lichter (2011), Garz (2014), Goidel and Langley (1995), Hofstetter and Dozier (1986), Kayser and Peress (2021), Nickl et al. (2025), Ryu (1982), Soroka (2016), and Soroka et al. (2018) and Soroka and Krupnikov (2021).

measures downstream effects on receivers' knowledge. In this setting, stronger attention incentives improve readability and engagement, but reduce information accuracy, revealing an attention-information trade-off. Her findings suggest that missing information, rather than explicit falsehoods, can distort beliefs among inattentive readers. The present study builds on these insights and broadens them along several dimensions. First, it moves from science summaries to more general news production by combining field data of several newspapers and topics with experiments with professional journalists as participants. I also isolate the emotional channel in headline choice, while Serra-Garcia (2025) studies broader shifts in linguistic style and overall informativeness. Second, the incentive treatments in this paper replicate central features of digital news markets by tying remuneration directly to realized reader behavior through pay-per-click and pay-per-subscription schemes, and by separately varying headline-level competition. In contrast, Serra-Garcia (2025) focuses on click and sharing incentives. Third, by linking the experimental effects to correlational patterns in real markets, the present study shows that exposure to digital competition is associated with more emotional and more negative headlines. Taken together, these elements broaden the evidence base from documentation of negativity bias and micro-level attention-information frictions to a more integrated view of how digital attention incentives shape headline choices and affect engagement and learning on the demand side.

Additional supply-side evidence comes from Balbuzanov et al. (2025), who demonstrate how performance-based pay shapes news supply in an actual newsroom environment: it increases page views, but reduces content diversity and positivity. Complementary evidence from Chopra et al. (2025) highlights an important demand-side distinction. In their setting, users click more on emotionally charged news when passively exposed to it, but actively choose neutral and fact-oriented versions when given control over tone. This gap between attention and preferences mirrors the present findings, in the sense that emotional content attracts engagement but does not necessarily benefit audiences. By combining large-scale text evidence with controlled experimental variation in both monetary and competitive attention incentives, and by tracing their consequences for a representative reader sample, this paper brings together these strands and illustrates how attention-driven markets shape emotional framing and its effects on engagement and information acquisition.

More broadly, the results contribute to work on how digitization transforms media markets and information environments (Acemoglu et al., 2025; Allcott et al., 2020; Boxell et al., 2017; Boxell et al., 2020; Bursztyn et al., 2024; Farrell, 2012; Gentzkow and J. M. Shapiro, 2011; Pariser, 2012). While much of this literature emphasizes polarization or echo chambers, the present study highlights a complementary mechanism. Digital attention incentives systematically amplify emotional and negative framing in news headlines, even in the absence of ideological motives. This mechanism also relates to evidence showing that variation in news sentiment shapes asset prices, risk perceptions, and macroeconomic beliefs (De Fiore et al., 2024; Tetlock, 2007). By identifying attention incentives as an upstream determinant of headline tone, the paper sheds light on a potential mechanism behind these documented effects.

The remainder of the paper is structured as follows. Section 2.2 presents the descriptive analysis of media content. Section 2.3 details the experimental variation in journalists' incentives and examines how monetary and competitive attention incentives shape headline choices. Section 2.4 analyzes how these changes affect readers' engagement and information acquisition. Section 2.5 integrates the findings and discusses their interpretation and validity. Section 2.6 concludes.

2.2 Descriptive Motivation: The Online-Offline Emotionality Gap

In this section, I compare headline sentiment in the online and print editions of major news outlets. The exercise is descriptive: it documents correlations rather than causal effects. To my knowledge, no study has systematically contrasted tonal differences between online and offline headlines over a longer time horizon, making this comparison a useful first step toward answering the broader research question. I begin with unconditional mean differences and then sequentially add outlet, time, and topic fixed effects, as well as controls for article length, agency content, and the sentiment of the full article text. These specifications reveal whether the observed gaps more pronounced in specific contexts. Finally, I match the subset of identical articles which have been published in both the online and the offline version of the newspapers and compare the average tonality of their headlines.[10]

2.2.1 Descriptive Data

The data on newspaper headlines analyzed in this part are obtained from the news aggregator *LexisNexis*, which systematically compiles articles, headlines, and metadata on a daily basis. For the primary analysis, I focus on economic news headlines published by the German outlets *BILD*, *Der Spiegel*, *Die Welt*, *Die Zeit*,

[10] Most newsrooms rely on the same editorial staff to produce articles for both formats, and the body text is often identical or very similar. Owing to space constraints, print versions may be shorter, and some online pieces never appear in print (and vice versa). Headlines, however, are frequently rewritten for the online or offline version.

and *Rheinische Post*. These outlets were chosen because their online and offline content are distinctly separated in the database, enabling a precise comparison. The selected newspapers differ notably in their political leanings, target audiences, and editorial style: for instance, *Die Welt* has a generally conservative orientation, while *Der Spiegel* is considered more liberal. *BILD* is a tabloid, whereas the others follow a traditional journalistic format. Additionally, *Rheinische Post* targets a regional audience, unlike the remaining outlets, which cater to national readership.

The only criterion applied when extracting articles was their classification by *LexisNexis* as economic news.[11] Thus, the sample covers the entire period for which both online and offline articles from each outlet are concurrently available. Alongside headlines, the dataset includes the publication date, medium (online or offline), article length (measured in units of 1,000 words), and full article text. Non-editorial material, such as obituaries, advertisements, and letters to the editor, was excluded.

Notably, the time periods covered differ substantially across outlets, restricting the common sample period available for all newspapers to late 2020 through mid-2022. Table 2.1 summarizes these differences in coverage and reports the number of available articles by outlet. The main analyses rely on the data on *all* outlets at all time-frames available. Appendix A.1.6.1 demonstrates that the key findings from this section hold robustly when analyzing most outlets individually, irrespective of variations in available time periods.

To evaluate the generalizability of the findings, I conduct an additional robustness analysis using headlines from non-economic news published by *BILD*, *Der Spiegel*, and from a non-German news outlet, namely *The New York Times*. These supplementary datasets comprise *all* online and offline headlines from each outlet within the time periods reported in the lower panel of Table 2.1. Consequently, the robustness check is not restricted to economic topics or the German news market. Unlike the main dataset, these robustness datasets do not include the full article texts; hence, it is not possible to apply text-based classification methods or match subsets at the article level. However, each headline in these datasets is accompanied by metadata on the news department ("ressort")[12] and, for offline articles, the respective page number of publication.

[11]The standard classification provided by *LexisNexis* is termed "Economic News and Economic Indicators."

[12]The ressort categorization broadly denotes the article's general topic, such as "sports," "business," or "culture."

Table 2.1: Description of Descriptive Datasets

Data for main analyses: Only economic news, includes article content

news outlet	**time-frame**	**N online**	**N offline**	**N total**
BILD	01/01/2017 - 01/06/2022	4,680	3,092	7,772
Der Spiegel	01/02/2003 - 01/06/2022	79,192	14,296	93,488
Die Welt	05/07/2009 - 01/06/2022	65,226	24,535	89,761
Die Zeit	10/01/2009 - 01/06/2022	50,643	64,260	114,903
Rheinische Post	10/09/2020 - 01/06/2022	11,348	22,593	33,941
all from above	all available points in time	211,089	128,776	339,865

Data for robustness checks: Includes all topics, but headlines only

news outlet	**time-frame**	**N online**	**N offline**	**N total**
BILD	01/01/2017 - 17/05/2022	66,720	97,576	164,296
Der Spiegel	01/01/2021 - 31/12/2021	19,122	4,896	24,018
The New York Times	01/01/2021 - 09/11/2021	25,230	29,216	54,446

Notes: Table 2.1 describes the timeframes and number of observations available for each news outlet that is part of the descriptive analyses. The data in for the main analyses contains articles on economic topics only, but comes with the entire content of the article. The datasets for the robustness checks news on all topics, but are limited to the headlines only.

2.2.2 Descriptive Analysis

I first classify the sentiment of each headline, generating the main variables of interest. These sentiment measures are then employed in regression analyses comparing the intensity of emotional language between the online and offline versions of each news outlet. Summary statistics for all variables included in these analyses are presented in Table A2.

A central conceptual choice for this concerns the definition of emotionality. In this paper, emotionality is defined as the absolute value of sentiment, where sentiment takes the values -1, 0, and 1 for negative, neutral, and positive headlines. This captures the presence of affective language independent of valence, reflecting that both positive and negative expressions convey emotion. More granular sentiment scales such as ranges from -2 to 2 are avoided, because human coders frequently disagree on distinctions between degrees of negativity or positivity, which reduces reliability and interpretability. This definition differs from semantic approaches such as Gennaro and Ash (2022), who locate words in a vector space along an emotion–reason dimension. In contrast, the present measure focuses on evaluative polarity and the extent to which headlines employ emotionally charged framing.

2.2.2.1 Sentiment Classification

Multiple methods exist to classify text sentiment as negative, positive, or neutral, and the literature has not converged on a single best practice. Instead, existing research highlights two important considerations: (i) the choice of classifier can significantly influence results, and (ii) different text domains require tailored classifiers[13] (Hartmann et al., 2022; A. H. Shapiro et al., 2022).

To identify the most suitable classification method for my dataset, I evaluate several algorithms against human-labeled sentiment categories derived from a randomly selected subset of headlines. Specifically, I compare three popular lexicon-based methods (the SentimentWortschatz dictionary (Remus et al., 2010), the Loughran-McDonald dictionary (Loughran and McDonald, 2011), and the Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto and Gilbert, 2014)) to a pre-trained transformer-based RoBERTa model (Y. Liu et al., 2019), a version of RoBERTa fine-tuned with another set of human classifications, and both pre-trained and fine-tuned versions of OpenAI's GPT (3.5 turbo) model.[14]

Descriptions of each classifier are provided in Appendix A.1.1. Among the considered approaches, the fine-tuned GPT model yields the highest accuracy and F1 score, and is thus selected for all subsequent analyses. Detailed performance metrics for all classifiers are reported in Table A1 in Appendix A.1.1.3. As a robustness check, Appendix A.1.6.4 demonstrates that the main results hold qualitatively when any of the alternative classification algorithms are employed.

2.2.2.2 Descriptive Comparison

The analysis compares two primary outcomes: *sentiment* and *emotionality*. For the *sentiment* measure, headlines classified as positive are assigned a value of 1, neutral headlines 0, and negative headlines -1.[15] However, this approach can lead to ambiguity at the aggregate level, as positive and negative headlines may offset each other. For instance, an outlet publishing entirely neutral content would receive the same average sentiment score as an outlet producing equal shares (50% each) of positive and negative headlines with no neutral content. To address this limitation, I also calculate an indicator of *emotionality*, defined as the absolute value of sentiment.

I compare the sentiment and emotionality of online versus offline headlines using linear regressions. To facilitate interpretation and effect-size comparisons, both

[13] For instance, dictionaries optimized for social media data may perform poorly when applied to financial text.

[14] Since only one algorithm is directly applicable to German text, I follow the recommendation of A. H. Shapiro et al., 2022 and translate the dataset into English using the Google Translate API prior to classification.

[15] This is a common method in the literature for summarizing the overall tonality of textual data.

outcomes are standardized to a mean of zero and a standard deviation of one.[16] Standard errors are bootstrapped at the news-outlet level to account for the small number of clusters (five outlets). The baseline regression specification, comparing online and offline headline tonality, is denoted in Equation 2.1.

$$tonality_i = \beta_0 + \beta_1 online_i + \epsilon_i \tag{2.1}$$

The dependent variable, $tonality_i$, represents either the standardized *sentiment* or *emotionality* of headline i. The main explanatory variable, $online_i$, is a dummy that equals one if the headline was published online and zero otherwise.

Any observed difference between online and offline headlines could arise from multiple underlying factors. For example, journalists may report more often on emotional topics online or frame given topics differently across the two media. Alternatively, specific emotional outlets or particular emotional time periods might drive the disparity. To examine these possibilities, I extend the baseline regression by sequentially adding fixed effects for outlet, topic of the article, and time, as well as controlling for the sentiment and emotionality of the full article text. Additionally, I estimate a comprehensive specification that simultaneously incorporates all these controls. The regression specifications and definitions of all included variables are provided in Appendix A.1.2.

Further, I construct matched subsets of the data, consisting of articles that appeared in both online and offline versions, to directly compare differences in sentiment and emotionality for identical articles across the two media. Specifically, I first decompose articles into overlapping five-word phrases, enabling efficient identification of similar texts through a hashing algorithm. Candidate article pairs flagged as highly similar by this algorithm are then re-ranked using semantic similarity scores computed via sentence-level embeddings and cosine similarity. I retain only pairs exceeding predefined similarity thresholds (cosine-similarity above 0.97, 0.95, 0.92, 0.90 or 0.85), resulting in five matched subsets with matched article pairs. A detailed description of the matching procedure is denoted in Appendix A.1.3.

2.2.3 Descriptive Results

2.2.3.1 Plain Online-Offline Comparison

Emotionality As depicted in the histogram in Figure 2.1b the largest share of the headlines on economic issues are classified as neutral. A comparison between online and offline versions reveals that online headlines are more frequently emotional,

[16] This standardization removes direct probabilistic interpretations of emotionality. Therefore, I replicate all descriptive analyses using the original, non-standardized measures in Appendix A.1.5.

meaning they are classified as either positive or negative. This difference is mainly driven by a higher share of negative headlines. A comparison of mean emotionality scores shows that online headlines are 18 percentage points more likely to be emotional than offline headlines ($p < 0.001$). This corresponds to a difference of 0.37 standard deviations. The respective means, together with 95 percent confidence intervals, are shown in Figure 2.1a.

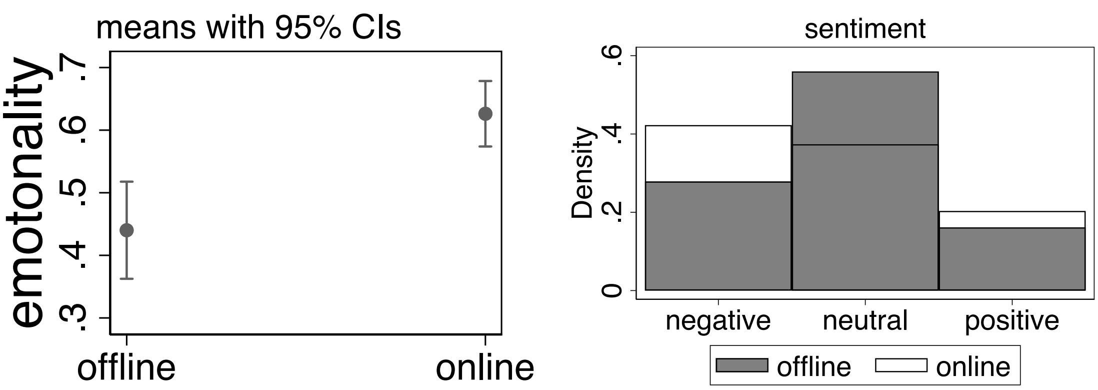
Sentiment In terms of sentiment, the data show that online headlines are on average more negative than offline ones. The magnitude of this difference, however, is considerably smaller: offline headlines are on average 0.14 standard deviations less negative than online headlines ($p < 0.001$). This smaller effect is likely at least partly due to the construction of the sentiment measure, as the higher shares of both positive and negative headlines partially offset each other when comparing mean sentiment values. Table 2.3 reports regression results.

2.2.3.2 Influence of Article Characteristics

To assess to which degree the differences in emotionality and sentiment are driven by specific article characteristics, I include various control variables in the regressions. The results show that the emotionality gap between online and offline headlines becomes somewhat smaller when controlling for content tonality, article topic and time fixed effects, as well as article length and the use of news agency material in the articles.

Figure 2.1: Descriptive Results: Emotionality and Sentiment of Headlines

(a) Means and 95%-confidence intervals **(b)** Distribution



Notes: Figure (a) illustrates the means and 95-percent confidence intervals of the emotionality for the headlines online and offline. The estimates displayed here are obtained by running the baseline regression as described in equation 2.1 (no controls). Figure (b) depicts the distribution of the sentiment measure. Values for the offline headlines are shaded in gray.

Table 2.2: OLS Estimates – Emotionality of Headlines in Standard Deviations

	(1)	(2)	(3)	(4)	(5)
online	0.3745***	0.2478***	0.3104***	0.3185***	0.2099***
	(0.0778)	(0.0463)	(0.0727)	(0.0594)	(0.0404)
content emotionality		0.6438***			0.5987***
		(0.0284)			(0.0225)
article length					0.0203
					(0.0176)
agency content					0.0412*
					(0.0245)
topic FE	no	no	yes	no	yes
time FE	no	no	no	yes	yes
Constant	-0.2326***	-0.5233***	-0.2632**	0.000	-0.0543*
	(0.0797)	(0.0197)	(0.1220)	(0.0385)	(0.0295)
R^2	0.0330	0.1306	0.0677	0.0171	0.1311
Observations	339,865	339,865	339,865	339,865	339,865

Notes: Table 2.2 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The offline headlines are always the reference group and the difference is expressed in standard deviations. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. The same regressions with the independent variable expressed as emotionality dummy are available in Appendix A.1.5. *** p<0.01, ** p<0.05, * p<0.1

Specifically, controlling for the emotionality of the content reduces the gap to 12 percentage points (0.25 standard deviations, $p < 0.001$). Including topic fixed effects lowers the estimate to 15 percentage points (0.31 standard deviations, $p < 0.001$), while adding time controls results in a difference of 16 percentage points (0.32 standard deviations, $p < 0.001$). Overall, these variables all contribute to the emotionality gap, yet none fully accounts for it. When all controls are included jointly, along with article length and agency content, the average difference in the emotionality of headlines declines to 10 percentage points (0.21 standard deviations, $p < 0.001$).

For sentiment, the difference between online and offline headlines becomes smaller once content sentiment and time fixed effects are controlled for, but it is only marginally affected by the other variables. When all controls are included jointly the average difference in the sentiment declines from 0.14 to 0.10 standard deviations. Table 2.2 reports the regression results for emotionality, and Table 2.3 presents those for sentiment.

2.2.3.3 Robustness

I assess the robustness of the main findings along several dimensions: time, outlet, topic, news market and classification methodology.

First, it is important to examine whether the observed differences are driven by specific outlets or particular time periods. To explore this, I plot the evolution of sentiment and emotionality over time, first for the entire dataset and then separately for each outlet. The results show that, for almost all outlets, online headlines are more emotional than offline headlines at all points in time, and that this difference remains relatively stable. Hence, the overall average difference is not driven by isolated periods with unusually large gaps. The pattern for sentiment is somewhat more nuanced. Although online headlines are on average more negative across outlets, this does not hold consistently for every outlet and every period and no clear time trend is detectable. A more detailed analysis is provided in Appendix A.1.6.1.

Second, I compare the sentiment and emotionality of headlines for identical articles appearing in both the online and offline versions of the same newspapers. Across all matched subsets, online headlines are on average more negative than their

Table 2.3: OLS Estimates – Sentiment of Headlines in Standard Deviations

	(1)	(2)	(3)	(4)	(5)
online	-0.1411***	-0.0954***	-0.1440***	-0.1287***	-0.1049***
	(0.0483)	(0.0197)	(0.0377)	(0.0353)	(0.0166)
content sentiment		0.7350***			0.7217***
		(0.0329)			(0.0309)
article length					-0.0308
					(0.0300)
agency content					0.0085
					(0.0265)
topic FE	no	no	yes	no	yes
time FE	no	no	no	yes	yes
Constant	0.0877	0.2721***	0.1086	0.0000	0.0271
	(0.0535)	(0.0419)	(0.0728)	(0.0233)	(0.0226)
R^2	0.0047	0.2688	0.0219	0.0028	0.2687
Observations	339,865	339,865	339,865	339,865	339,865

Notes: Table 2.3 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The offline headlines are always the reference group and the difference is expressed in standard deviations. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. The same regressions with the independent variable expressed as sentiment ranging from -1 to 1 are available in Appendix A.1.5. *** p<0.01, ** p<0.05, * p<0.1

offline counterparts. With respect to emotionality, the point estimates indicate that online headlines are also slightly more emotional, although this difference is not statistically significant. The exact matching procedure is described in Appendix A.1.3. Detailed results for the matched datasets are presented in Appendix A.1.6.3.

Third, I examine the results using all previously discussed sentiment classification algorithms. Appendix A.1.6.4 shows that the main findings are consistent across all classifiers considered in Section 2.2.2.1: regardless of the algorithm used, online headlines are on average more emotional and more negative.

Forth, I test whether the results are specific to economic topics. This does not appear to be the case, as the main findings are replicated in a two samples where all topics of an outlet are included (for example also culture, politics, sports, etc.) over a shorter time frame. Detailed descriptions of these additional datasets and regression results are available Appendix A.1.6.5.

Fifth, I assess whether the difference in emotionality is specific to the German news market, by running the same analysis with headlines from the *New York Times*, where a similar gap in emotionality and sentiment appears. Further details are provided in Appendix A.1.6.5.

Overall, the most robust finding from the descriptive analysis is that online headlines are on average more emotional, especially in the direction of negativity, than offline headlines.

> **Result 1: Headlines of news outlets are written more emotionally, and in particular more negatively, for online audiences than for offline audiences.**

The descriptive patterns presented above reveal a consistent empirical regularity: across outlets, topics, and time, online headlines are more emotional and, on average, more negative than those in print. These correlations, however, do not by themselves identify an underlying mechanism. The tonal differences could reflect journalists' incentives, differences in audience composition, or other factors. To move beyond correlations, subsection 2.2.4 provides suggestive evidence pointing to intensified attention incentives for journalists as a key mechanism. Section 2.3 then provides clear causal evidence from randomized experiments.

2.2.4 Suggestive Evidence: Attention Incentives and Headline Emotionality

Directly observing changes in journalists' attention incentives in practice is difficult. Explicit monetary reforms are rare, typically endogenous, and often unobservable to outside researchers. Instead, I examine whether the descriptive data exhibit patterns consistent with stronger attention incentives leading to more emotional headlines. I pursue two complementary empirical approaches. First, I exploit variation in the competitive environment in which headlines are displayed. When multiple headlines compete for readers' attention on the same page, the marginal value of attracting a click plausibly increases. I therefore test whether settings with greater headline-level competition are associated with higher headline emotionality. Second, I examine the introduction of digital paywalls. Paywalls change the economic environment of news production by tying revenues more directly to reader engagement and conversion. This shift plausibly strengthens incentives to attract attention at the headline stage, as headlines become a key gateway to subscription decisions. While paywall adoption is not exogenous, it provides a meaningful setting to assess whether increases in the returns to attention coincide with changes in headline tone.

2.2.4.1 Headline Competition in Offline Markets

Offline print markets offer a clear and observable gradient of headline-level competition: front-page headlines face substantially stronger competitive pressure than headlines printed inside the newspaper. Using the robustness datasets[17], I compare the emotionality and sentiment of front-page headlines to non-front-page headlines. Consistent with the idea that higher competition drives emotional headline framing, front-page headlines are found to be systematically more emotional and more negative on average. Across outlets, emotionality is between 0.07 and 0.26 standard deviations higher and sentiment between 0.02 and 0.18 standard deviations lower for front-page headlines. All differences are statistically significant at the one-percent level. Detailed estimates are reported in Appendix A.1.7.1.

2.2.4.2 Headline Competition in Online Markets

An additional consistency check comes from a temporary reduction in digital competition caused by a legal dispute between German publishers and news aggregators in 2014. During this episode, one outlet in the sample (*Die Welt*) was removed from several major aggregators after a legal dispute. Because aggregators prominently display competing headlines side-by-side, their removal plausibly reduced headline-level competitive pressure for this outlet. Thereby, this episode provides a rare opportunity to examine whether changes in competitive exposure are accompanied by tonal shifts that align with the experimental patterns. Following the approach

[17] The offline data from the main dataset cannot be used for this purpose, as it does not contain page numbers.

suggested by Meyer et al. (2024), I compare average emotionality in the 18 months before the debate about the ancillary copyright law (pre-March 2013) with the 18 months after the aggregators' removal (post-August 2014). Two simple comparisons are informative: (i) *Die Welt* relative to the two unaffected outlets and (ii) *Die Welt*'s online headlines relative to its own offline edition over the same period.

Both comparisons point in the same direction: Emotionality at *Welt Online* declines slightly after the removal, with back-of-the-envelope Diff-in-Diff estimates ranging from about 0.08 to 0.13 standard deviations depending on the control group. No meaningful changes appear for sentiment. These patterns replicate qualitatively across alternative classifiers and remain similar when expanding or reducing the time window. Given the extremely limited number of outlets and the absence of a credible identification strategy (for example, spillovers from online to offline versions are quite plausible), the evidence should be interpreted only as suggestive. Nevertheless, the direction of the shift is consistent with the idea that stronger attention pressure tends to increase emotional framing. Detailed results are provided in Appendix A.1.7.2.

2.2.4.3 Introduction of Paywalls

Three of the five outlets in the sample introduce a digital paywall during the period covered by the data. I therefore examine whether the gap in headline emotionality between online and print editions changes around these introductions. For two of the three outlets, the emotionality of online headlines increases relative to the print edition following the introduction of the paywall, while the third outlet shows little change. These patterns are consistent with the idea that paywalls strengthen attention incentives in the online environment. Appendix A.1.7.3 presents the full time series before and after a paywall introduction for all outlets.

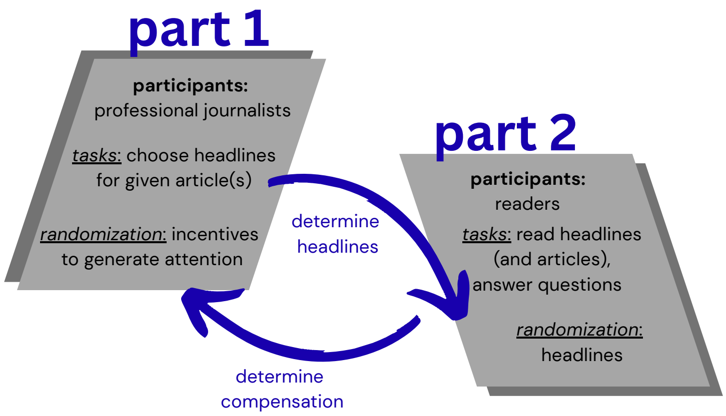
2.3 The Causal Effect of Attention Incentives on Headline Emotionality

This section examines whether stronger attention incentives *causally* shift how journalists frame news headlines. Building on the descriptive evidence in Section 2.2, I study how professional journalists respond when their remuneration depends on audience engagement rather than a fixed payment. Two experiments randomly vary the strength of these incentives, mirroring the contrast between digital and analogue media environments, and thereby provide direct evidence on how attention-based rewards shape headline choices.

2.3.1 Main Experiment with Professional Journalists

An online experiment with professional journalists ($N = 402$) was conducted between October 11 and November 30, 2024. Participation was restricted to individuals who were currently or had been employed as journalists within the previous twelve months. Recruitment took place via professional mailing lists of journalist associations and alumni networks of journalism programs at the Cologne School of Journalism and Dortmund University. The study was implemented using the survey software *Qualtrics*. The median completion time was 12.02 minutes. Participants received an average total payment of €13.60, which in the treatment groups depended partly on individual performance.

Figure 2.2: Overview Connection of Journalists and Readers Experiments



Notes: Figure 2.2 gives an overview of the two parts of the set of experiments. Part one is the journalist experiment and part 2 is the experiment with readers.

In addition to the experiment with journalists, a connected survey experiment with a sample of $N = 1,617$ "readers" was conducted. As illustrated in Figure 2.2, the behavior of the readers partially influenced the journalists' compensation and the behavior of the journalists' partially influenced the readers' choice set. Further details on the readers part of the experiment are provided in Section 2.4. Ethical approval for both experiments was granted by the Ethics Committee of the Faculty of Management, Economics and Social Sciences at the University of Cologne (reference: 240019LB). The study was pre-registered in the AEA RCT Registry under AEARCTR-0014049.[18] The design builds on and extends an earlier set of pilot experiments conducted in late 2021, which are described in detail in Appendix A.2.2 and A.3.2.

[18]The pre-registration is available at https://www.socialscienceregistry.org/trials/14049.

2.3.1.1 Experimental Design: Main Journalist Experiment

The journalist experiment is designed to study how incentive structures influence journalistic content choices, particularly the emotional tone of news headlines. Professional journalists were recruited as participants and randomly assigned to one of three incentive schemes that mimic central differences between offline and digital media markets. Figure 2.3 illustrates the experimental procedures. A transcript of the experimental instructions (translation to English) is provided in Appendix A.4.2. Participants were randomly assigned to one of three experimental conditions:

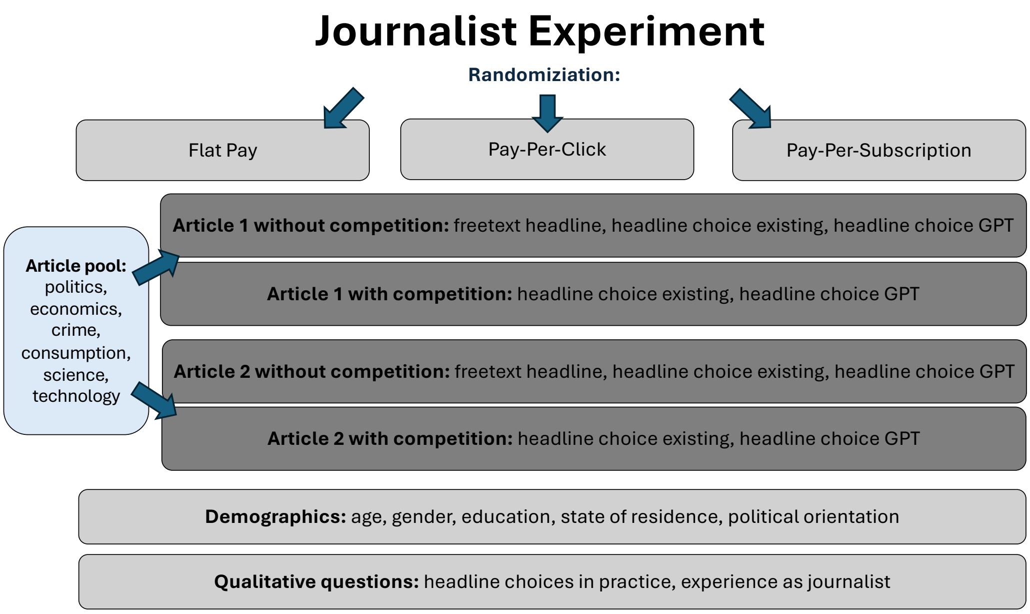
1. **Flat-Pay (Control Group):** Journalists in this group received a fixed compensation of €15 for completing the survey, independent of their headline choices.

2. **Pay-per-Click (Treatment Group I):** Participants received a base payment of €6 plus a performance-dependent bonus linked to the number of reader clicks generated by the headlines they selected. For each selected headline, earnings equaled €4 multiplied by the share of readers (from a representative reader sample) who had previously clicked on that headline. In addition, authors of the five self-written headlines generating the highest click rates among readers received an extra €50 bonus.

3. **Pay-per-Subscription (Treatment Group II)):** Participants received a base payment of €6 plus a performance-dependent bonus linked to the share of readers who subscribed to a newsletter after viewing their selected headlines. The bonus was calculated as €8 multiplied by the proportion of readers who subscribed. As in the click condition, the authors of the five self-written headlines producing the highest number of newsletter subscriptions earned an additional €50.

The pay-per-click condition approximates advertising-driven digital markets in which journalists are rewarded for maximizing immediate audience engagement, while the pay-per-subscription condition mirrors reader-revenue models that reward the generation of deeper, sustained interest among readers. The flat-pay condition provides a benchmark representing traditional journalism without such strong audience incentives. Journalists in all experimental conditions were informed that their choices had been or would be displayed to a group of readers which is representative for the German population. After assignment to the groups, which happened without participants' notice, all journalists completed a standardized sequence of headline tasks and follow-up questions.

Figure 2.3: Overview Experimental Procedures



Notes: Figure 2.3 gives an overview of the experimental procedures. Randomization is indicated with arrows. The decisions generating the main outcomes are shaded in grey.

In addition to the between-subject variation in incentive schemes, the experimental design also introduces a within-subject variation in competitive context of the headlines. Each journalist made headline choices both in a non-competitive setting, where the headline choice was evaluated by readers in isolation, and in a competitive setting, where the same article appeared alongside a competing headline from an article about a different aspect of the same topic[19]. This manipulation captures a central difference between offline and digital news environments: Whether a headline competes directly for attention. Implementing this variation within subjects allows for an estimate of the effect of competition on headline tone while maintaining statistical power, as the comparison does not rely on cross-participant differences and the generation of six between-subject groups would have made it unfeasible to recruit enough participants for each group. The use of a single competing headline per trial ensures a controlled representation of audience competition in online settings, without introducing excessive noise from multiple simultaneous comparisons. This

[19]The competing headline was always a neutrally formulated headline which was fully disclosed to journalists. The full set of all competing headlines is denoted in Appendix A.2.1.1.

design feature allows to directly test whether journalists respond to competitive pressure by selecting more emotional headlines.

Each participant completed headline tasks for two randomly drawn short news articles covering two topics from the areas of politics, economics, technology, crime, science, and consumer issues. Articles were selected from recent news agency reports by *dpa* and were at most one month old at the time of the experiment. To ensure comparability across tasks, the article pool was constructed using several criteria: articles had to allow for headlines with different tonalities, contain at least one numerical fact, and be associated with at least one positive, neutral, or negative headline used by real newspapers. The resulting articles are characterized by similar length and readability. The composition of the article pool is described in detail in Appendix A.2.1.1. The headline tasks proceeded in three steps:

1. **Headline Generation:** Participants read a short article and wrote an original headline in a free-text field.

2. **Headline Choice Existing:** Participants chose one headline from a set of alternatives that had been used by existing newspapers reporting on the agency article at hand and that varied systematically in sentiment: One option was positive, one neutral and the third or negative. For illustration, one set of existing headlines presented to participants read as follows: *Positive:* "Inflation near ECB target — path cleared for interest rate cuts." *Neutral:* "Inflation in the euro area falls to 2.2 percent." *Negative:* "ECB target in sight — but central bankers urge caution."

3. **Headline Choice GPT:** Participants chose one headline from a set of alternatives that had been generated by a large language model (OpenAi's ChatGPT o3 turbo) and that varied only in sentiment: One option was positive, one neutral and the third or negative[20].

All choice sets are denoted in Appendix A.2.1.1. The three headline tasks span a spectrum between realism and experimental control. Writing a headline in free text most closely resembles real newsroom practice, but introduces substantial variation, which can reduce statistical power in finite samples. The two headline choice tasks therefore impose progressively more structure. Choosing among existing headlines preserves realistic phrasing while limiting variation, whereas the GPT-generated headline choice task holds all dimensions constant except sentiment, providing the highest degree of experimental control. Together, the tasks allow treatment effects

[20] Human coders validated that this classification worked well.

to be examined both in a realistic setting and in environments with lower noise and greater statistical power.

Task 2 and task 3 were by each participant completed first in an uncompetitive context and then in an competitive context. In the competitive context the same article as in the uncompetitive context was shown, but this time participants chose among the same three alternatives while they are informed that their choice was presented alongside a competing headline about a different article on a related topic. This simulates environments where multiple stories vie simultaneously for readers' attention, such as online news feeds.

A separate reader experiment (described in Section 2.4) was conducted to determine the performance-based payments in the incentive treatments and to measure how headline sentiment affects reader behavior. In that study, a representative German sample was exposed to the headlines generated by journalists and their engagement (clicks and subscriptions) was recorded.

2.3.1.2 Definition of Outcome Variables

The main outcomes are the *emotionality* and *sentiment* of journalists' headlines choices across tasks. Consistent with the definitions introduced in Section 2.2, each headline is classified as positive, neutral, or negative using the fine-tuned supervised machine-learning classifier described there. *Emotionality* is defined as a binary variable equal to 1 if a headline is classified as either positive or negative, and 0 otherwise. *Sentiment* is measured on a scale ranging from -1 (negative) to +1 (positive). A subset of classifications was validated by independent human coders to confirm the reliability of the automated labeling.

Additional dependent variables include differences between competitive and non-competitive contexts (within-subject variation) and a free-text description in which journalists describe their perceived influences on headline choice. At the end of the experiment, participants provided demographic and professional information including age, gender, education, region (East/West Germany), type of media outlet (print, online, radio, TV), job role (e.g., reporter, editor, columnist) and their political orientation. These variables are used as controls in the regression analyses.

The substantial variation in journalists' self-written headlines permits analysis of additional headline attributes. In particular, I examine whether headlines contain factual inaccuracies relative to the underlying article.[21] Two independent human coders evaluate each headline using the full text of the corresponding article. They classify a binary indicator for factual mistakes that equals one if the headline contains

[21] By design, all options in the existing-headline and GPT-headline choice sets are factually correct.

a concrete error relative to the article, such as incorrect directions or magnitudes of change, incorrect locations, incorrect counts, or causal claims not supported by the article, and zero otherwise. In contrast to the other outcomes, this measure was not part of the pre-registration and should therefore be interpreted as exploratory.

2.3.1.3 Sample Descriptives: Main Journalist Experiment

The experiment involved 402 journalists randomly assigned to one of three payment schemes. Participants were diverse in terms of age, gender, professional background, and political orientation. On average, they were around 35 years old. 44 percent of participants identified as male and 61 percent held an editorial position. A large share (about two thirds) worked for newspapers or magazines, while others were employed in broadcasting or digital-only outlets. Most participants had a university degree and were located in Western Germany, reflecting the general composition of the German journalist population. Figures A14 and A15 in Appendix A.2.1.3 illustrate the distributions of age and political orientation in the sample.

Table A28 reports a balance check across treatment groups at the participant level. The randomization produced groups that are overall similar in observable characteristics. None of the core demographic variables (age, gender, education, or region) differ significantly across treatments. Small differences appear for some subcategories of political preference and outlet type: for instance, journalists identifying with the left party (*Die Linke*) are somewhat more common in the flat-pay group, while print journalists are slightly overrepresented there relative to the pay-per-click group. Given the number of comparisons, such differences are within the range expected by chance. To account for these minor imbalances, all main regressions include controls for demographics, political orientation, education, and outlet type. The estimated treatment effects remain virtually unchanged or become even larger when these controls are added, confirming that the results are not driven by composition differences across groups.

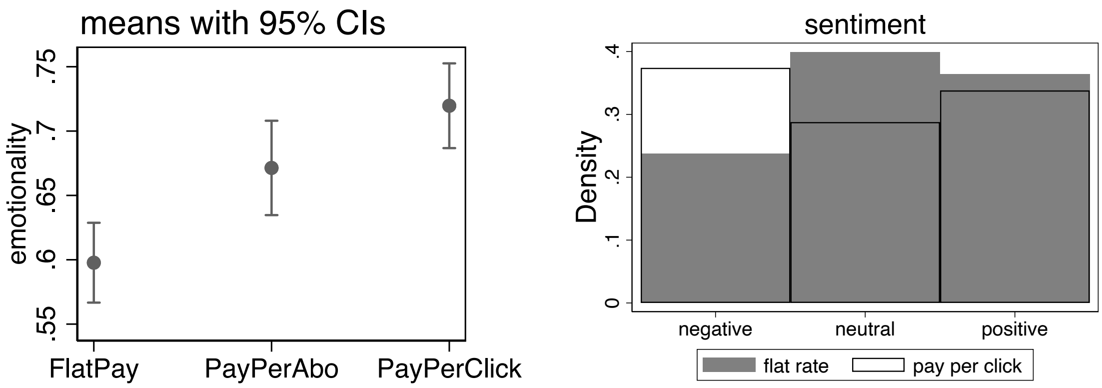
Although the experiment draws on a broad cross-section of journalists, participation is voluntary and the sample is thereby not a probability sample of the full German journalist population. The composition of the sample however aligns relatively closely with the most recent representative survey of German journalists (Loosen et al., 2023). Compared to this benchmark, the sample is somewhat more female (45 percent male vs. 55 percent male in the representative survey) and younger on average (mean age 34 vs. 45). Educational attainment is nearly identical: 68 percent of participants in the experiment hold a university degree, compared to 67 percent in the representative sample. The distribution of outlet types also broadly matches the population of journalists. In the experiment, 60.2 percent work for newspapers or magazines (59.4 percent in Loosen et al. 2023), and 23.6 percent for broadcasters (33.9 percent in the representative data), with the difference primarily reflecting a somewhat larger share of journalists from digital-only or alternative online outlets in the experimental sample.

2.3.1.4 Results: Main Journalist Experiment

Pooled headline choice Table 2.4 reports the average treatment effects on headline choices. Here, all choices of each are considered and standard errors are clustered at the level of the individual journalist. Compared to the flat-rate baseline, both performance-based payment schemes lead to significantly more negative and more emotional headlines. Under pay-per-click incentives, average sentiment decreases by about 0.17 points (corresponding to an effect size of 0.22 standard deviations in the control group), while emotionality increases by roughly 12 percentage points (0.25 standard deviations). Similar effects are observed for the pay-per-subscription scheme. For emotionality, they are slightly smaller and for sentiment slightly larger. The inclusion of controls leaves the results virtually unchanged, indicating that the treatment effects are not driven by differences in observable journalist characteristics or by the topic of the specific article. The differences in means and the distributions of sentiment classes are illustrated in figure 2.4. Table A14 in Appendix A.2.1.4 displays the main results with all control variables in detail.

Figure 2.4: Experimental Results: Emotionality and Sentiment of Headline Choices

(a) Means and 95%-confidence intervals **(b)** Distribution



Notes: Figure 2.4(a) illustrates the means and 95-percent confidence intervals of the emotionality for the headlines choices of journalists in the different experimental conditions. The estimates displayed here are obtained by running comparative regressions as in Table 2.4 controlling for the standard set of covariates. Figure 2.4(b) depicts the distribution of the sentiment measure across the two most extreme treatment arms (FlatPay and PayPerClick). Values for the FlatPay condition are shaded in gray.

Table 2.4: OLS Estimates – ATE on Headline Choice

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	-0.1623***	-0.1706***	0.1109***	0.1220***
	(0.0445)	(0.0412)	(0.0251)	(0.0233)
Pay-per-abo	-0.1980***	-0.2092***	0.0737**	0.0737***
	(0.0483)	(0.0452)	(0.0267)	(0.0248)
controls	no	yes	no	yes
Constant	0.1265***	-0.2429	0.6013***	0.9201***
	(0.0257)	(0.1685)	(0.0171)	(0.1006)
R^2	0.0118	0.0719	0.0101	0.0947
Observations	4,020	4,020	4,020	4,020

Notes: Table 2.4 reports OLS estimates with robust standard errors clustered at the level of the individual journalist (402 clusters) in parentheses. The flat-rate (fixed pay) group is the reference category. Columns (2) and (4) include controls: age, male, location, political orientation, newsroom type, editor status, education and topic of the article. *** p<0.01, ** p<0.05, * p<0.1.

Result 2: Incentives to generate attention induce journalists to select emotional, and particularly more negative, headlines.

Free-text headlines Table A15 in Appendix A.2.1.4 presents the results for the subset of self-written headlines (free-text field). Both performance-based payment schemes significantly decrease sentiment, indicating that journalists formulate more negatively framed headlines when compensation depends on audience engagement. Effects on emotionality are positive and become statistically significant once controls and topic fixed effects are included, suggesting that performance incentives also increase the use of emotionally charged language in this subset. Taken together, these results imply that incentive-induced shifts in tone extend beyond the selection among predefined options to journalists' own wording choices.

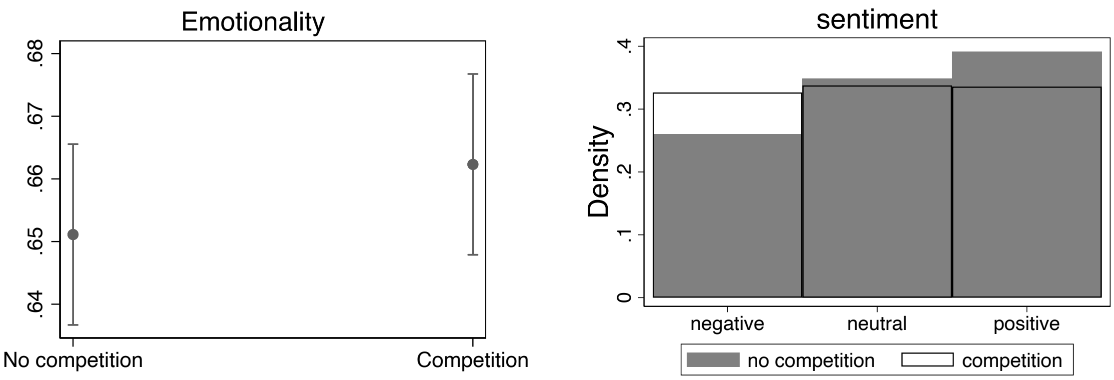
Existing headlines Table A16 presents the results for the subset of existing headlines that journalists selected. They had the choice between a positive, neutral and negative headline that came from actual news sources. These tasks provide more structure than the free-text condition, as journalists choose from a set of authentic options, yet they still allow for substantially greater variation than the later machine-generated headlines. Both performance-based payment schemes again lead to more negative and more emotional headline choices, with effect sizes comparable to those observed in the pooled sample. Under both pay-per-click and pay-per-subscription incentives, sentiment decreases by roughly 0.18 to 0.30 points (statistically significant across all comparisons), while emotionality increases by about 6 to 10 percentage points (only statistically significant for the comparisons of the FlatPay and the Pay-Per-Click group, while the estimates for the Pay-Per-Abo group consistently lie between those two values). The inclusion of controls leaves the results virtually unchanged.

GPT-generated headlines Table A17 reports the results for tasks with GPT-generated headline options. These tasks offer the most structure in the experiment and therefore less residual variation than free-text or existing-headline choices. Both performance-based payment schemes again shift selections toward more negative and more emotional headlines, with magnitudes that are comparable to those in the pooled sample. Under pay-per-click, sentiment decreases by about 0.13 to 0.21 points (0.18 to 0.28 standard deviations in the control group) and emotionality rises by roughly 9 to 23 percentage points (0.18 to 0.46 standard deviations) across specifications. Under pay-per-subscription, sentiment decreases by about 0.11 to 0.25 points (0.15 to 0.34 standard deviations) and emotionality increases by 7 to 11 percentage points (0.14 to 0.24 standard deviations). The inclusion of controls leaves the estimates largely unchanged.

Within subjects: Competition Table A20 reports the between-subject effects of

Figure 2.5: Experimental Results: Within-subject variation in competition

(a) Emotionality **(b)** Sentiment



Notes: Figure 2.5(a) illustrates the means and 95-percent confidence intervals of the *emotionality* for the headlines choices of journalists in the conditions with the different degrees of competition (within-subject). The estimates displayed here are obtained by running comparative regressions as in Table A20. Figure 2.4(b) illustrates the distribution of *sentiment* for the headlines choices of journalists in the conditions with the different degrees of competition (within-subject).

competition on headline choices. Across all tasks, competition significantly decreases the average sentiment of selected headlines, while leaving emotionality unaffected. In the fixed-effects specification, the presence of competition reduces sentiment by approximately 0.12 points (corresponding to about 0.17 standard deviations in the control group), indicating a systematic shift toward more negatively framed headlines when journalists receive direct competition on the headline level[22]. The estimated effect on emotionality is positive but statistically insignificant. Figure 2.5 illustrates the differences in means and distribution in terms of sentiment classes.

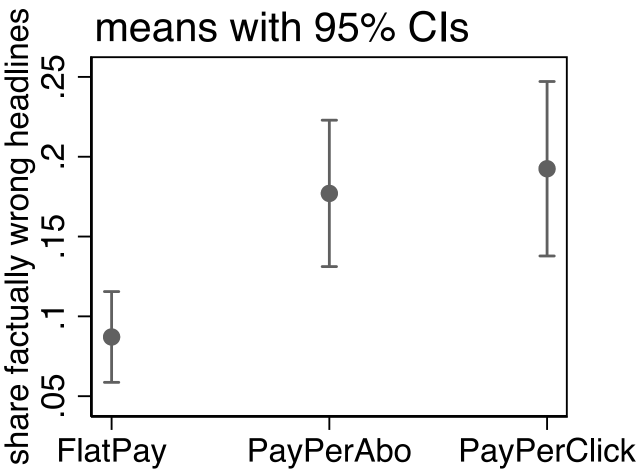
To test whether this effects differs across incentive schemes, Appendix Table A21 reports regressions including interaction terms between competition and the two performance-based pay treatments. The results indicate that neither the pay-per-click nor the pay-per-subscription incentives significantly moderate the impact of competition. In both cases, the interaction terms are small and statistically insignificant, and a joint Wald test confirms that the effect of competition does not differ across pay schemes ($p = 0.86$).

Qualitative Answers Open-ended responses on headline selection criteria were analyzed using a word-frequency approach. After removing stop-words, the forty most frequently mentioned terms were visualized in a word cloud (Appendix Figure A16). The resulting pattern aligns closely with the experimental findings: Frequent references to *arousal*, *reader*, *attention*, and *sensationalism* suggest that journalists prioritize audience-oriented and attention-grabbing headline features.

Exploratory Analysis: Factual Accuracy of Headlines In addition to the pre-registered outcomes discussed above, I explore whether performance incentives affect the factual accuracy of journalists' self-written headlines. Prior research documents that factual errors in journalistic reporting are common (L. M. Berger et al., 2026; Cacciatore, 2021; Dempster et al., 2022; Oxman et al., 2022; Sumner et al., 2016). A preliminary inspection of the free-text headlines revealed several potential factual inaccuracies. For example, one headline reads "Life is finally getting cheaper again – inflation is falling!", although the underlying article reports only a decline in the inflation rate rather than falling prices.

[22]The within-subject design always presents non-competitive trials before competitive ones, so the contrast could in principle be confounded with trial-order effects. Two checks suggest otherwise: (i) relative to a uniform-random-choice benchmark (mean sentiment = 0), observed sentiment averages +0.131 without competition and +0.009 with competition, so competition moves journalists from a positively tilted distribution toward random rather than into the negative; (ii) within each block (where competition is held constant), choices drift toward more positive sentiment over trial positions, biasing the within-subject estimate *against* finding a negative effect and making the reported coefficient a conservative lower bound.

Figure 2.6: Exploratory Result: Factual Mistakes in Freetext Headlines



Notes: Figure 2.6 illustrates the means and 95-percent confidence intervals of the share of factually incorrect headlines of journalists in the different experimental conditions. The estimates displayed here are obtained by running comparative regressions as in Table A22 (including controls).

Two independent human coders therefore systematically evaluated each headline relative to the corresponding article and classified whether it contained a factual mistake, such as incorrect directions or magnitudes of change, incorrect counts or locations, or causal claims not supported by the article text. Inter-coder agreement is high: the two coders agree in 87.2 percent of cases ($\kappa = 0.48$, $p < 0.001$). The outcome variable used in the analysis is the average of the two coders' classifications.

Table A22 and Figure 2.6 summarize the results. In the control group, approximately 8.4 percent of self-written headlines contain a factual mistake according to the coders' average classification. Both performance-based payment schemes significantly increase the likelihood of producing a factually incorrect headline. The probability of a headline containing a factual mistake rises by 11.5 percentage points in the PayPerClick treatment ($p = 0.001$) and by 9.1 percentage points in the PayPerAbo treatment ($p = 0.001$). These estimates remain similar and statistically significant when including the full set of controls. These exploratory findings suggest that stronger attention incentives may not only affect the emotional tone of headlines but also reduce their factual precision.

2.3.1.5 Heterogeneity of Journalist Reactions

To explore whether responsiveness to performance-based pay varies across demographic subgroups, Tables A18 and A19 report specifications that interact treatment indicators with observable characteristics. Overall, heterogeneity remains limited. Age and education do not systematically moderate the effects, and there is no robust evidence of differential effects on sentiment across subgroups. One consistent pattern emerges for emotionality: journalists identifying as left-leaning show a weaker increase under pay-per-subscription (negative and statistically significant interaction in columns (5)–(6) of Table A18). Taken together, these results suggest only modest subgroup variation, with left ideology dampening the subscription-incentives-induced rise in emotionality, while effects on sentiment remain broadly similar across groups.

Appendix Table A23 reports treatment effects by article topic. The results again show some variation in point estimates across content domains, though the differences are not statistically significant. Incentive effects are most pronounced for politics, science, and crime, where both performance-based payment schemes lead to more negative and more emotional headlines. Estimated effects for economics, technology, and consumption are smaller in magnitude and not statistically significant. The absence of significance likely reflects limited statistical power, as each journalist evaluated only two of the six articles, resulting in smaller topic-specific samples. Joint Wald tests of equality across topics yield $F(5, 401) = 0.33$ ($p = 0.89$) for the pay-per-click interactions and $F(5, 401) = 0.17$ ($p = 0.98$) for the pay-per-subscription interactions, suggesting that treatment effects do not differ significantly across topics.

2.3.1.6 Robustness of Journalist Reactions

Multiple Hypothesis Testing As robustness check, I apply a Bonferroni correction across the two pre-registered primary outcomes (sentiment and emotionality). All four treatment effects remain statistically significant even under this conservative adjustment (Appendix Table A24); the adjusted p-values range from 0.000025 (PayPerClick–emotionality) to 0.012 (PayPerAbo–emotionality).

Human Validation of Sentiment Classifications To validate the automated classification of headline tone, I compare the model-based measures of emotionality and sentiment with independent human coding. A human coder evaluated the same set of free-text headlines and classified both dimensions using the same definitions. Agreement between the model and human coding is substantial. For emotionality, the two measures agree in 74.8% of cases ($\kappa = 0.42$, $p < 0.001$). For sentiment, agreement amounts to 71.1% ($\kappa = 0.56$, $p < 0.001$). These levels of agreement are consistent with moderate to substantial inter-rater reliability and suggest that the automated measures capture headline tone in a manner broadly aligned with human judgment. If the human classifications are used as outcome measures treatment effects become larger both in magnitude and statistical significance.

Alternative Classifier Reclassifying the free-text headlines with the RoBERTa model introduced in Section 2.2 produces treatment effects that broadly point in the same qualitative direction as those obtained with the GPT-based classifier. However, the estimates are markedly less precise and lose statistical significance, reflecting the substantially smaller free-text sample (Appendix Table A27).

Spillover Check To rule out potential spillovers from the within-subject structure, I re-estimate all treatment effects using only the first headline decision, which was always made in the non-competitive context. The results (Appendix Table A25) are nearly identical to the main estimates: both performance-based incentive schemes significantly increase emotionality and decrease sentiment, and the effect sizes remain highly stable across specifications with and without controls.

Alternative Clustering I further re-estimate the main specifications using two-way clustered standard errors (journalist $\times$ topic) to allow for arbitrary dependence across both participants and article topics. Appendix Table A26 shows that the estimated treatment effects are virtually unchanged in significance.

2.3.2 Earlier Experiment with Journalists

A prior online experiment with professional journalists ($N = 201$; conducted in late 2021) provides consistent evidence for the attention incentives mechanism documented above. Journalists were randomly assigned to a flat-pay control or a pay-per-click scheme whose remuneration was tied to click rates generated in a separate reader sample ($N = 299$). Given a neutral wire-service article and three factually correct headline options (positive/neutral/negative), pay-per-click increased the probability of choosing an emotional headline by about 20 percent (0.41 SD; p = 0.001), driven by more frequent selection of both the positive and the negative option, while average sentiment did not change because these choices offset each other. The direction and structure of effects mirror the main experiment: Stronger attention incentives shift selection toward more emotional headlines. This strengths the external validity of the main findings. Design details and additional results of this first experiment are provided in Appendix A.2.2.

2.4 Reader Responses to Attention Incentives

The preceding analysis has documented systematic shifts in headline tone that originate from digital attention incentives. The next step is to examine how these shifts affect the reception of information by readers. Section 2.4 analyzes the audience side of the market using controlled reader experiments, which test whether exposure to emotional and competitive news environments influences readers' engagement, factual learning, opinion polarization and reported mood.

2.4.1 Main Reader Experiment

An online experiment with a sample of $N = 1{,}617$ readers was conducted between October 11 and October 25, 2024. Participants were recruited via the market research company *Bilendi* and were representative of the German population in terms of age, gender, and federal state. The experiment was directly linked to the journalist study described above and used the headlines from the existing and GPT-generated headline pools as experimental stimuli[23]. It was implemented using the survey software *Qualtrics*. The median completion time was 8 minutes and 50 seconds. Participants received a fixed remuneration for completing the study, with the payment rate determined by the market research company. To keep the setting as close as possible to reading behavior in real news environments, none of the outcome measures was monetarily incentivized. Ethical approval was granted by the Ethics Committee of the Faculty of Management, Economics and Social Sciences at the University of Cologne (reference: 240019LB). The study was pre-registered alongside the journalist experiment in the AEA RCT Registry under AEARCTR-0014049.[24]

2.4.1.1 Experimental Design: Main Reader Experiment

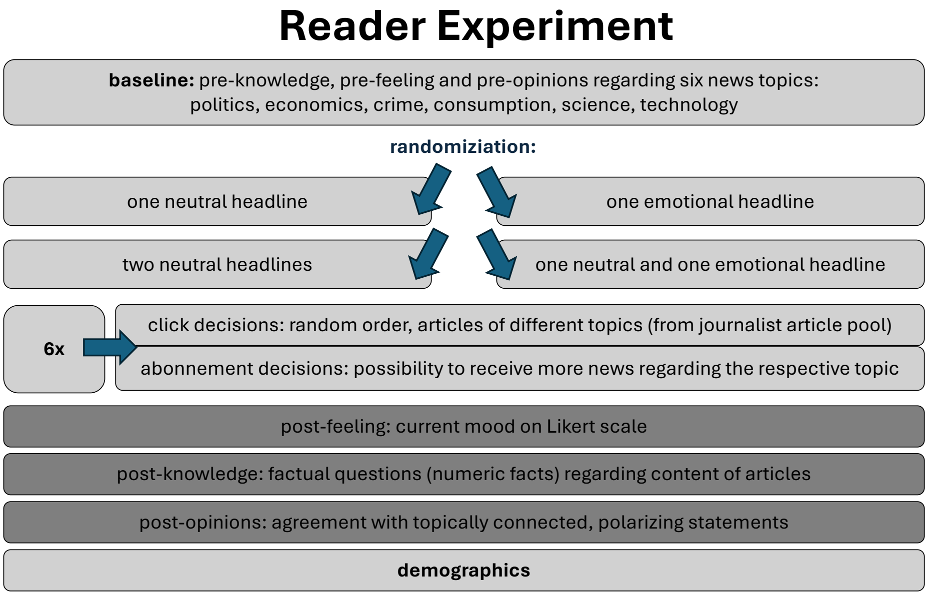
In the experiment, participants were randomly assigned to one of four main treatment groups that varied in the emotional tone and competitive setting of the presented headlines: (1) neutral headlines without competition, (2) neutral headlines with competition, (3) emotional headlines without competition, and (4) emotional headlines with competition. Within the emotional treatments, half of the participants were exposed to positive and half to negative headlines. Each participant viewed six news headlines and could decide whether to click on one or several of the presented headlines, read the associated article, and optionally subscribe to a topic-specific newsletter. The experiment further elicited factual knowledge, opinion polarization, and affective states both before and after exposure to the news content. Figure 2.7 provides an overview of the experimental procedures.

The survey followed a structured sequence consisting of five main parts. First, participants reported their current affective state ("pre-feeling") on a 0–100 scale. Second, they answered baseline questions on topical knowledge and opinions related

[23]The journalists incentives for the freetext headlines were determined by a separate survey, which is described in Section 2.4.2.

[24]The pre-registration is available at https://www.socialscienceregistry.org/trials/14049.

Figure 2.7: Overview of Procedures: Main Readers Experiment



Notes: Figure 2.7 gives an overview of the experimental procedures. Randomization is indicated with arrows. The main outcomes are shaded in grey.

to the six thematic domains covered in the subsequent articles (politics, technology, science, the economy, crime, and consumer products). Third, participants entered the exposure phase, in which they were presented with a series of six news headlines drawn from the experimental headline pool. Depending on their treatment assignment, participants saw either one or two headlines per topic. Between experimental groups, the sentiment of these headlines and the competitive setting varied.

Participants could click on any headline to read the corresponding article and were then offered the option to subscribe to a topic-specific newsletter. In the fourth part, participants again reported their affective state ("post-feeling") and answered factual knowledge questions, as well as the same set of opinion statements to measure belief accuracy and opinion polarization. Finally, participants provided feedback on the study and, if applicable, entered their email address to receive the selected newsletters. The experimental instructions are provided in Appendix A.4.4.

2.4.1.2 Definition of Outcome Variables

Most outcome variables are analyzed at the headline level, with standard errors clustered at the participant level. This section details the construction of the main outcome measures.

Click dummy A dummy variable equaling 1 if a headline was clicked on and 0 otherwise.

Subscription dummy A dummy variable equaling 1 if a reader chose to subcribe to a newsletter after being exposed to a specific headline and 0 otherwise.

Factual knowledge Factual knowledge is measured based on participants' accuracy in answering six numerical estimation questions at the end of the survey. These questions refer to specific numerical values that are mentioned in the articles. For each question, the (absolute) deviation from the correct value is computed, where smaller deviations indicate higher factual accuracy. Let $y_{i\ell}$ denote the participant's answer and y^*_ℓ the correct value for topic ℓ. The main outcome is the logarithm of the absolute deviation,

$$d_{i\ell} = \ln |\, y_{i\ell} - y^*_\ell \,|,$$

which allows for a more symmetric distribution and downweights large deviations that otherwise dominate the mean. Each $d_{i\ell}$ is winsorized at the 95th percentile to reduce the influence of extreme values and then standardized to have mean zero and unit variance.[25]

Time in news environment This variable measures the time in seconds that a participant spends on the decision page for a given topic, which contains one headline in the no-competition conditions and two headlines in the competition conditions, including any time spent on the linked article(s) after a click. To limit the influence of extreme values (for example when participants leave the page open and return later), time in the news environment is winsorized at the 95th percentile.[26]

Polarization of opinions Opinion polarization is measured based on participants' responses to six attitudinal statements, one for each article topic. Each item is

[25]Since ln(0) is undefined, observations with zero deviation ($y_{i\ell} = y^*_\ell$) are set to zero after the log transformation. Winsorizing instead at the 90th or 99th percentile leaves results qualitatively unchanged. Results are robust to using the absolute deviation without log transformation instead; see Appendix Table A34.

[26]Winsorizing instead at the 90th or 99th percentile leaves results qualitatively unchanged.

recorded on a five-point scale ranging from –2 (“strongly disagree”) to +2 (“strongly agree”). For every topic, the absolute value of the post-exposure response is taken so that higher values indicate stronger agreement or disagreement and thus greater polarization.

Mood Mood is measured using participants’ self-reported affective state after exposure to the experimental stimuli on a continuous scale from 0 (“very bad”) to 100 (“very good”).

2.4.1.3 Statistical Analysis: Main Reader Experiment

For the empirical analysis baseline regressions first compare main outcomes across treatment groups. Additionally, an instrumental variables (IV) approach provides complementary estimates for outcomes that require active article consumption, capturing effects among participants induced to read the article by the treatment.

Baseline estimation The baseline specification compares outcomes between treatment groups according to

$$Y_i = \beta_0 + \beta_1 \mathrm{TG}_i + \beta_2 \mathbf{X}_i + \varepsilon_i, \tag{2.2}$$

where Y_i denotes the outcome for headline i, TG_i indicates treatment assignment of the participant exposed to the headline, and $\mathbf{X}_i$ is a vector of covariates (age, gender, income, occupation, political orientation, education, news consumption, prior knowledge, attitudes, mood and article topic).

The main comparison is between the Emotional/Competition (`EMO/C`) and Neutral/No-Competition (`NEU/NC`) groups, representing the highest and lowest exposure to attention-driven market features. Additional pairwise comparisons separately identify effects of emotionality (`EMO/C` vs. `NEU/C`; `EMO/NC` vs. `NEU/NC`) and competition (`EMO/C` vs. `EMO/NC`; `NEU/C` vs. `NEU/NC`).

The regression is estimated at the headline level, with six observations per participant. Standard errors are clustered by participant to account for within-subject correlation.

Analysis of Mood For the analysis of mood, the emotionality treatment is further split into positive and negative headline conditions, allowing for a three-way comparison (positive, neutral, negative). This distinction captures the plausibly asymmetric effects of positive and negative headlines on participants’ emotional states.

IV Specification For the outcome factual knowledge, treatment effects can only fully materialize once participants choose to click on a headline and read at least part of the article. Hence, estimates from the baseline specification should be interpreted as intention-to-treat (ITT) effects. To examine how knowledge changes when readers actually engage with the content, an instrumental variables (IV) approach is employed. The endogenous click variable is instrumented with the exogenous assignment to treatment. First-stage results confirming instrument relevance are reported in Appendix A.3.1.1.

2.4.1.4 Sample Descriptives: Main Reader Experiment

The experiment includes 1,617 readers, evenly distributed across the four treatment arms. Participants are representative of the German population in terms of gender, age, and region of residence. Table A31 reports highly similar means across conditions for all pre-treatment covariates. Only few pairwise differences reach conventional significance levels (see Table A32), consistent with random variation given the large number of tests. No systematic imbalances are observed, suggesting that randomization was successful. To account for the few existing imbalances, the full set of descriptive variables is included as controls in the main regression specifications.

2.4.1.5 Results: Main Reader Experiment

Click rates Table A42 shows that attention-driven features substantially increase readers' propensity to click. Emotional headlines under competition raise click rates by about 18 to 20 percentage points relative to neutral headlines without competition ($p < 0.01$). Emotionality alone slightly raises clicking under competition but lowers it in the absence of competition, while competition increases the clicking probability regardless of headline tone.

Subscription rates Table A43 presents results for newsletter subscriptions. Here, effects are modest and mostly negative. Relative to the neutral, no-competition condition, emotional headlines under competition reduce subscription rates by around 3 percentage points ($p = 0.03$), with similar declines when comparing emotional to non-emotional content within the competition arm.

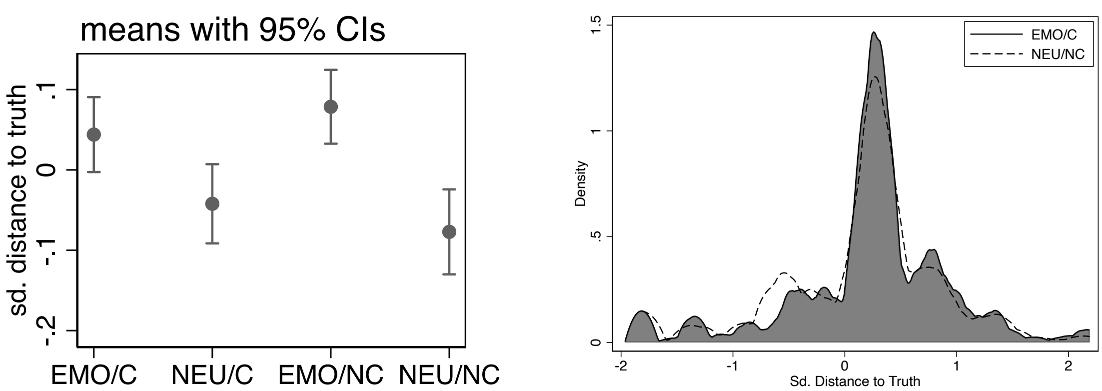
Factual knowledge Table A33 reports intention-to-treat (ITT) effects on the standardized log distance to the truth, where higher values indicate lower factual accuracy. These estimates capture the effect of being assigned to different headline environments, regardless of whether participants ultimately clicked on the article.

Readers assigned to the *Emotionality with Competition* condition perform significantly worse on the knowledge questions than those in the *Neutral without Competition* condition, with knowledge declining on average by 0.10–0.12 standard deviations ($p < 0.01$). For example, in one of the knowledge questions asking readers to estimate the core inflation rate in the euro area, participants assigned to the emotional headline environment misestimate the true value by 0.92 percentage points more on average ($p < 0.01$). Similar deteriorations arise when comparing emotional to neutral headlines within the competitive arm (0.08–0.10 SD; $p \leq 0.05$) and within the noncompetitive arm (0.13–0.16 SD; $p < 0.01$). Competition by itself does not affect factual knowledge (columns 7–10). Figure 2.8 illustrates these patterns: emotional headline environments shift the distribution toward larger deviations from the truth. Results are robust to alternative outcome definitions using the non-log standardized distance to truth (see Appendix Table A34).

To examine effects among readers who actually engaged with the content, Table A35 presents instrumental-variable estimates using treatment assignment as an instrument for clicking. The first-stage results (Table A30) show that assignment strongly predicts clicking behavior. For readers induced to click on emotional articles under competition, factual accuracy decreases substantially (0.5–0.7 SD, $p \leq 0.05$). In the questions about core inflation, readers induced to click on the article mistake the true value on average by 2.96 percentage points ($p < 0.01$). In contrast, for emotional headlines without competition (where assignment slightly reduces the probability of

Figure 2.8: Experimental Results: Factual Knowledge of Readers

(a) Means and 95%-confidence intervals **(b)** Kernel Densities



Notes: Figure 2.8(a) illustrates the means and 95-percent confidence intervals of the standardized distance to the truth of readers in the different experimental conditions. The estimates displayed here are obtained by running comparative regressions as in Table A33 controlling for the standard set of covariates. Figure 2.8(b) depicts the kernel densities of the knowledge measure across the two main treatment arms. Values for the Emotionality/Competition condition are shaded in gray.

clicking) the negative first stage implies that the negative IV coefficient (-1.9 SD, $p < 0.01$) reflects that the subset of readers who still clicked on the emotional article performed substantially better on the factual knowledge questions. This estimate should be interpreted with caution given the weaker first stage ($F = 7.6$).

Overall, the results suggest that emotional headlines reduce factual learning in competitive environments. In noncompetitive settings, the negative ITT effect appears largely driven by selection into reading: emotional headlines discourage some readers from clicking, while the subset who still engage performs relatively well on the knowledge questions.

Time in news environment Table A36 shows that exposure to *Emotionality with Competition* increases the total time participants spend in the news environment by about 8.6–8.9 seconds relative to *Neutral without Competition* ($p < 0.01$). Emotionality alone does not meaningfully change this outcome within the competition or no-competition settings (columns 3–6), whereas competition by itself lengthens time in the news environment by roughly 7.4–9.8 seconds for both emotional and neutral headlines (columns 7–10; $p < 0.01$). Results remain robust when using alternative winsorization thresholds (Tables A37 and A38).

Because this measure captures the total time spent in the news environment for a given topic, it naturally combines time spent viewing headlines with any time spent reading linked articles. In the competition conditions, participants are shown two headlines instead of one, which increases both the amount of visible content and the number of possible click paths. Supplementary analyses that normalize the measure by the number of displayed headlines[27] or that condition on the number of clicks show that exposure time per headline or per article is similar across treatments. This suggests that part of the competition effect on total time reflects the additional content presented on the page.

Polarization of opinions Table A39 shows no detectable treatment effects on opinion polarization. Point estimates are small across all contrasts and remain statistically indistinguishable from zero with or without controls. Table A40 examines absolute changes in opinions between the post and pre measures. Estimates are near zero in the main and emotionality versus neutral contrasts, and not statistically significant. The only clear pattern appears within the emotional arm, where competition is associated with a modest reduction in absolute opinion change of about 0.09 to 0.10 units (columns 7–8; $p < 0.05$). No corresponding effect is observed for neutral headlines. Overall, there is no evidence that emotionality increases polarization

[27] That is, dividing the time measure by two in competition conditions.

or amplifies opinion shifts; if anything, competition slightly dampens within-arm movement for emotional headlines.

Mood Table A41 reports post-treatment mood evaluations. Across all comparisons, effects are small and mostly statistically insignificant. Neither positive nor negative emotional headlines systematically alter readers' mood relative to neutral content once covariates are included. The only notable result appears within the emotional arm, where competition lowers mood ratings by about 4.8 points for positive headlines ($p < 0.05$), while no corresponding effect is observed for negative ones. Prior feelings are, as expected, the strongest predictor of post-treatment mood, accounting for nearly all explained variation ($R^2 \approx 0.9$). Overall, headline tone and competitive framing exert limited influence on readers' emotional state.

The evidence presented above shows that attention-driven environments systematically shape how audiences process news. Emotional headlines, especially when combined with competition, reduce factual retention, while competition increases the likelihood that readers enter an article and the time spent with news content.

> **Result 3: Attention driven news environments increase readers' engagement, but at the same time reduce their factual knowledge.**

2.4.1.6 Heterogeneity of Reader Results

To explore whether responses to attention incentives differ across subgroups, the baseline specification (Equation C.1) is extended by interacting the treatment indicator with pre-treatment characteristics. Each regression is estimated on the main comparison sample between the Emotional/Competition (`EMO/C`) and Neutral/No-Competition (`NEU/NC`) conditions, with standard errors clustered at the participant level (see Table A44).

Demographics Reader responses exhibit limited heterogeneity across standard demographics. Interactions with gender and age are small and statistically indistinguishable from zero across knowledge, reading time, polarization, mood, clicks, and newsletter sign-ups. Political orientation shows no systematic moderation either.

Prior Knowledge and Opinions A first set of heterogeneity results concerns the information status and prior beliefs of readers. Readers with strong prior opinions spend more time on headlines shown under attention incentives (2.647* seconds) and exhibit slightly larger polarization responses (0.106^{*} on a scale from 0 to 2). Low knowledge readers have a higher baseline error rate, but the incremental knowledge reduction from attention incentives is smaller for them (-0.165^{**} standard deviations).

Mood A second dimension of heterogeneity relates to the affective state of readers. A positive pre-treatment mood dampens attention: Among readers in a good pre-mood, the increase in time in news environment induced by attention incentives is substantially smaller (-6.096^{**} seconds). Education also moderates affective responses. The small negative mood effect observed under attention incentives is attenuated for highly educated readers (*post mood*, `NEG/C`$\times$*HighEdu*$= 2.930^{**}$ higher mood on a scale from 0 to 100).

Media Use The data further suggests that media consumption habits matter for downstream engagement. High digital readers are less likely to request the newsletter when exposed to attention incentives (-0.070^{**} percentage points), indicating that frequent users of digital news respond to the same incentives with lower subscription intent.

Headline type: Existing vs. GPT In the reader experiment participants are randomly assigned to either existing headlines or GPT generated counterparts. Interacting the treatment with a headline type indicator shows that the effects of attention incentives are largely similar across the two formats, except for the following exceptions.

First, attention incentives attract more time in news environment for GPT generated headlines than for existing ones. The interaction is positive and statistically significant (2.73^{**} seconds), which indicates that GPT generated emotional framing captures somewhat more attention.

Second, the overall null effect on polarization masks heterogeneity. For GPT generated headlines, attention incentives do not shift polarization (0.020, $p = 0.587$). For existing headlines, attention incentives reduce polarization modestly, as reflected in the negative and statistically significant interaction term (-0.096^{*} on a scale from 0 to 2). Thus, existing headlines exhibit a slight depolarizing response, while GPT generated headlines leave polarization unchanged.

Across all other outcomes, including mood, clicking, and newsletter sign ups, differences between existing and GPT generated headlines remain small and statistically insignificant.

Overall, heterogeneity is modest relative to the average effects. Given the large number of comparisons, the described subgroup patterns should be read as suggestive and interpreted cautiously.

2.4.1.7 Robustness of Reader Results

Multiple Hypothesis Testing After adjusting for multiple testing using a Bonferroni correction over the four primary, pre-registered outcomes (knowledge, time in news environment, mood, and opinion polarization), the effects on knowledge and time in news environment remain statistically significant. See Appendix Table A45 for raw and adjusted p-values (knowledge: $p_{\text{Bonf.}} = 0.0036$; time in news environment: $p_{\text{Bonf.}} < 0.0001$).

Order effects Table A46 examines whether treatment effects vary with the position of a topic within the sequence of six articles. While readers in the `EMO/C` condition spend significantly more time on an article and show lower factual knowledge on average, the interaction terms indicate no systematic moderation by order for knowledge, polarization, or mood. The only notable pattern emerges for time in news environment, where the treatment effect diminishes for later articles, suggesting mild fatigue rather than order dependent treatment heterogeneity.

Alternative Clustering Results remain nearly identical when clustering standard errors two-way at the participant and topic level. The effects on knowledge, time in news environment and clicking are estimated slightly less precise, but remain statistically significant at conventional levels.

2.4.2 Evaluation of Freetext Headlines

A follow-up survey with a smaller, non-representative *Prolific* sample was conducted from November 15 to 19, 2024, after completion of the journalist experiment. Its primary purpose was to collect click and newsletter subscription data for the journalists' self-written headlines to determine bonus eligibility[28]. Participants had to estimate how likely they were to click on a certain headline or subscribe to a related newsletter after reading this headline on a five-point Likert scale.

The fine-tuned classifier from Section 2.2 is used to classify headline sentiment, and correlations between emotionality, sentiment, and reader responses are examined.

[28] Including these items in the main reader experiment was not feasible because the large number of self-written headlines would have required an impractically large sample to maintain clean comparisons across treatment groups. In addition, the reader experiment had to be completed before the journalist experiment so that main payments could be processed promptly.

In contrast to clicking behavior in the more realistic (and cleanly randomized), larger reader experiment, participants stated probability of clicking on a emotional headline is negatively correlated with emotionality ($p = 0.048$) and positively correlated with sentiment ($p = 0.001$). This pattern of stated reader preferences is consistent with findings by Chopra et al., 2025, who document that users click more on emotionally charged news when passively exposed to it, but actively choose neutral and fact-oriented versions when allowed to adjust tone.

2.4.3 Earlier Experiment with Readers

A complementary online experiment with readers ($N = 299$; conducted in late 2021, contemporaneous with the first journalist study) provides audience-side evidence consistent with the main findings above. Participants were randomly assigned to one of three factually correct headlines (positive, neutral, negative) for the same wire-service article used in the journalist task. They could click to read the article at a small cost of €0.05, then reported current mood and completed incentivized belief-elicitation tasks on GDP growth and the DAX.

Emotional framing shifted affect and expectations but left clicking and investment choices unchanged. The positive headline improved current mood by about 0.2 standard deviations (p=0.001), while emotional headlines increased the absolute forecast error for GDP in 2022 by roughly 0.23 standard deviations (p=0.047). Negative headlines reduced DAX expectations by about 0.27–0.31 standard deviations ($p \leq 0.05$). Treatment effects on investment and click rates were small and statistically indistinguishable from zero.

Taken together, the pilot aligns closely with the larger study. Measuring accuracy as the absolute distance between participants' expectations and the benchmark forecast yields the same pattern as the knowledge outcome in the main experiment: Exposure to emotional headlines increases errors. This concordance strengthens the external validity of the central finding that emotional framing reduces factual accuracy among readers. Design details of the first study and full results are provided in Appendix A.4.3 and Appendix Tables A48–A50.

2.5 Discussion

2.5.1 Joint Interpretation and Alternative Channels

The results together reveal a consistent mechanism linking digital incentives to information supply and reception. Stronger competition for attention induces more emotional and more negative headlines, and such framing in turn raises audience engagement while lowering factual learning. These findings indicate that shifts in market incentives can systematically alter not only journalistic choices, but also how information is processed by audiences.

A natural question is how much of the observed online–offline gap in emotionality in the descriptive data this mechanism can account for. In the descriptive data, the share of emotional headlines is roughly 18 percentage points higher online than in print, and about 10 percentage points higher after controlling for content tone, time, a set of article characteristics and topic composition. A simple back-of-the-envelope comparison suggests that the experimental incentive effects are of comparable magnitude. Introducing click-based incentives increases the probability that journalists write an emotional headline by about 12 percentage points and subscription incentives by about 7 percentage points. While these figures should not be interpreted as a structural decomposition of the online–offline gap, they indicate that attention-based incentives alone are large enough to generate a substantial share of the tonal differences observed between online and print news.

At the same time, the online–offline gap may reflect the cumulative influence of numerous features of digital markets: platform design, algorithmic ranking, real-time analytics, social-media sharing, editorial workflows, and selection into online and offline readership. These different technological features of digital information environments should not be seen as alternative or competing channels, but rather as extensions of the same fundamental incentive mechanism: Algorithmic recommendation systems, social-media visibility, and real-time feedback loops amplify the returns to emotional and especially negative content. For example, recent large-scale evidence shows that negative news is almost twice as likely to be shared on social media as neutral or positive news (Watson et al., 2024). Such amplification increases the potential audience for emotionally framed content and thereby strengthens the incentives for journalists to produce it, regardless of whether their compensation is explicitly tied to engagement.

While audience composition may also differ between online and offline contexts, the reader experiment suggests that behavioral responses to emotional headlines are remarkably stable across demographic groups and habitual online/offline readers. This limits the scope for purely demand-based explanations and underscores that supply-side incentives, broadly construed, are central to understanding the tonal divergence.

2.5.2 External Validity

The experimental settings deliberately abstract from many complexities of real-world journalism. Professional participants faced simplified headline-selection tasks and short time horizons, whereas newsroom environments involve continuous feedback, editorial coordination, and repeated interactions with audiences and platforms. Likewise, the reader experiments capture short-term effects of limited exposure. In practice, repeated encounters with emotional headlines or sustained competition for attention may lead to larger and more persistent shifts in engagement, trust, and belief formation. In addition, the reader experiment isolates the informational consequences of headline framing alone. Exploratory evidence from the journalist experiment suggests that stronger attention incentives may also increase the incidence of factual mistakes in self-written headlines. To the extent that such inaccuracies affect how readers interpret the underlying articles, the knowledge effects measured here likely represent a lower bound of the broader informational distortions that may arise in attention-driven news environments.

Several features of the design, however, enhance the external relevance of the findings. First, the experimental variation mirrors a central economic feature of digital news markets: journalists increasingly observe granular performance metrics and operate in environments where visibility and professional evaluation depend on measurable engagement. While explicit pay-per-click contracts remain uncommon, real-world incentives are shaped by similar feedback loops, for example through analytics dashboards, social media traffic, employer expectations, and platform algorithms. These forces constitute different institutional manifestations of the same underlying mechanism identified in the experiment.

Second, the descriptive evidence provides an important real-world benchmark. The tonal gap between online and offline headlines closely aligns with the direction and structure of the treatment effects, suggesting that differences in incentive strength of the magnitude induced in the experiment can plausibly contribute to the observed divergence. External platform-level research further shows that digital environments disproportionately reward emotional and negative content (Watson et al., 2024), reinforcing the broader incentive gradient journalists face.

Third, field evidence from professional newsrooms is consistent with the identified mechanism. Balbuzanov et al. (2025) find that linking compensation of freelancers to clicks increases negative and political content while reducing informational diversity. This closely parallels the experimental responses to stronger attention incentives. On the demand side, Chopra et al. (2025) show in large-scale, longer-run interventions that emotional and negative framing reliably boosts engagement, confirming that the behavioral patterns observed in the reader experiment extend to realistic settings.

These elements suggest that although stylized, the experiments presented in this paper capture a core incentive mechanism operating in digital markets. The external evidence indicates that this mechanism is active in the field and amplified by the technological design of online platforms. As such, the results offer a credible causal explanation for patterns observed at scale, even as further work is needed to study long-run effects and newsroom-level dynamics in detail.

2.5.3 Industry Relevance and Broader Implications

The results have implications for how media organizations and policymakers may think about digital news environments. The experimental design distinguishes click-based and subscription-based pay, but in practice many newsrooms combine both. The evidence suggests partial complementarity rather than a clean substitution. Subscription incentives also increase emotional headline selection, albeit less strongly than click incentives, consistent with outlets facing competition for initial attention even when long-run value depends on credibility. A plausible editorial response is dynamic: use emotional framing to attract initial attention, while shifting toward accuracy and completeness within the article to preserve retention and trust. Quantifying this dynamic trade-off between acquisition and retention incentives remains an important avenue for future research.

From a macroeconomic perspective, these dynamics matter for central banks and other institutions whose communication relies on accurate and balanced media coverage. Overly emotional framing of economic news can distort public expectations, complicate forecast calibration, and amplify cycles of optimism and pessimism. Incorporating an understanding of these incentive-driven distortions into communication strategies and forecasting models may help policymakers anticipate public reactions and design messages that remain informative even in highly competitive digital environments.

Technological design choices also offer potential levers for mitigating excessive emotional amplification. Recent evidence on AI-based news customization (Chopra et al., 2025) shows that when users receive greater control over presentation features such as tone or complexity, many opt for more neutral, fact-oriented formats. Such tools cannot eliminate market incentives, but they may help realign the supply of digital news with underlying user preferences, thereby counteracting some of the emotional amplification generated by competitive attention markets. Incorporating mechanisms of this kind into editorial and platform design could thus complement traditional editorial and regulatory approaches.

2.6 Conclusion

This paper combines large-scale descriptive evidence with experiments involving professional journalists and a representative reader sample to study how digital attention incentives shape the tone and informational value of news headlines. Descriptively, online headlines are substantially more emotional and more negative than their offline counterparts. The experiments show that strengthening attention incentives, both through performance-based pay and through direct headline-level competition, leads journalists to select more emotional and on average more negative headlines for identical underlying articles, and that such framing increases engagement but reduces factual learning among readers. These results support a coherent attention-information mechanism through which digital attention incentives influence both the supply of news and the way audiences process it.

The findings complement existing work on media bias by highlighting a market-based channel that operates independently of ideological motives or ownership interests. Exploratory analyses of self-written headlines further suggest that stronger attention incentives may affect additional dimensions of information quality, in particular factual precision, pointing to a broader set of trade-offs that arise in attention-driven environments. As news consumption continues to migrate online, understanding these incentive distortions becomes crucial for evaluating how effectively the media fulfills its informational role in modern democracies.

Data Statement

All four experiments were pre-registered in the AEA RCT Registry under AEARCTR-0008658 (2021 pilots) and AEARCTR-0014049 (2024 main experiments). All experiments were approved by the Ethics Committee of the Faculty of Management, Economics and Social Sciences at the University of Cologne (references 210036LM for the 2021 pilots and 240019LB for the 2024 main experiments). The experiments received financial support from the Center for Social and Economic Behavior (C-SEB) at the University of Cologne and the Joachim Herz Stiftung. The headline sample was retrieved from *LexisNexis* under the University of Cologne's research subscription. *LexisNexis*'s terms of service preclude redistribution of the raw text. The replication package therefore contains the processed analysis dataset together with the full processing code. The processed descriptive dataset, anonymised participant-level data from all four experiments, the fine-tuned GPT-3.5 classifier (model identifier and training data), all analysis code are deposited at https://doi.org/10.17605/OSF.IO/XD325.

Chapter 3

Debunking “fake news” on social media: immediate and short-term effects of fact-checking and media literacy interventions

Paper Information

Authors: Lara Marie Berger, Dr. Anna Kerkhof, Felix Mindl, Prof. Dr. Johannes Münster

Status: Published in Journal of Public Economics, Volume 245, 2025, 105345.

Abstract

We conduct a randomized survey experiment to compare the immediate and short-term effects of fact-checking to a brief media literacy intervention. We show that fact-checking primarily affects the specific fake news it directly addresses, whereas media literacy helps to distinguish between false and correct information more generally, both immediately and around two weeks after the intervention. A plausible mechanism is that media literacy enables participants to critically evaluate social media postings, while fact-checking fails to enhance their skills as much. Our results promote media literacy as an effective tool to fight fake news, that is cheap, scalable, and easy-to-implement.

3.1 Introduction

The emergence and spread of "fake news" – i.e., false or misleading information presented as news – has led to widespread concerns (e.g., D. M. Lazer et al., 2018). Social media like Facebook and X (formerly Twitter) are especially prone catalysts for the evolution of fake news and have consequently come to the fore of public and academic debates. Indeed, recent evidence suggests that 50% of users who see fake news on social media say that they believe them (Allcott and Gentzkow, 2017).

What helps users to distinguish between false and correct information on social media? Policymakers support fact-checkers on the one hand, and media literacy initiatives on the other.[1] Independent fact-checking organizations complement such campaigns, and large social media have started to implement AI-based procedures that flag suspicious content, too (Allen et al., 2021). Yet, it is unclear whether these remedies function as desired: empirical evidence on the effectiveness of fact-checking is mixed (Jerit and Zhao, 2020; Vraga and Bode, 2017), knowledge on the impact of media literacy is scarce (A. M. Guess et al., 2020), and a direct comparison of the effect of these interventions on the beliefs and attitudes of consumers does not exist at all.[2]

We address this gap with a large-scale randomized survey experiment on the immediate and short-term effects of fact-checking and media literacy interventions. In the experiment, we expose a sample of German residents to false and correct statements on health-related topics – Corona vaccines and nutrition – that we retrieve from the German version of Facebook ("fakes" and "facts").[3] One group of participants receives additional fact-checks that debunk some of the fakes explicitly. Another group gets ten "Tips to spot fake news" before exposure to the fakes and facts as a brief media literacy intervention. Then, we compare the two treatment groups to participants who do not receive an intervention, which informs us about the immediate impact of our interventions. To study the interventions' efficacy in the short-run, we survey the same participants around two weeks later in a second, similar experimental wave without any additional interventions.

[1] See, e.g., `https://digital-strategy.ec.europa.eu/en/library/disinformation-threat-democracy-brochure` and `https://digital-strategy.ec.europa.eu/en/policies/online-disinformation` (viewed August 2022) for further information on efforts by the European Union.

[2] Guriev et al. (2023) compare the effect of an attention nudge (which could be thought of as a minimalistic media literacy enhancement) to fact-checking and other policies with sharing of fake posts as an outcome.

[3] Online misinformation about health can have severe consequences for quality of life and even for mortality risks (see Swire-Thompson and D. Lazer (2020) for a review of the literature on online misinformation about health, and Allen et al. (2024) for a study of vaccine related misinformation on Facebook).

Our results demonstrate that the effectiveness of fact-checking tends to be limited to the fakes that are being corrected, whereas media literacy helps to distinguish between fakes and facts more generally, both immediately and about two weeks after the intervention. A plausible explanation is that the media literacy intervention raises participants' attention and enables them to critically evaluate the postings' accuracy. Fact-checking, in contrast, turns participants into passive recipients of the specific corrections and thus fails to markedly enhance their skills.[4]

Specifically, we consider three main outcomes: the perceived credibility of fakes and facts, factual knowledge on the topics discussed therein, and attitudes towards Corona vaccination and dietary supplements (the fakes on nutrition promote the consumption of needless protein and vitamin preparations). The idea is to study a coherent cognitive chain: Do the interventions reduce the perceived credibility of fakes (but not of facts)? If yes, does that translate into better factual knowledge? If yes, does this entail a change in attitudes?

We find that both interventions reduce the credibility of fakes on Corona vaccines (which are corrected by fact-checks) immediately, but only the media literacy intervention reduces the credibility of fakes on nutrition (which are not corrected by fact-checks), both immediately and in the short-run (two weeks later). Moreover, both interventions improve participants' factual knowledge immediately, but only the media literacy intervention in the short-run. Finally, while the media literacy intervention raises participants' willingness to get vaccinated (or boostered) against Covid-19 both immediately and in the short-run, fact-checking has no such effect. Crucially, neither intervention reduces the credibility of facts or factual knowledge on the topics discussed therein, i.e., participants do not become more skeptical towards social media postings per se. Hence, in an environment where not every posting can be fact-checked, media literacy interventions are likely to be more effective than fact-checking on average.

Our subgroup analyses reveal that participants who are well informed from the beginning are less likely to benefit from the interventions than participants whose prior beliefs are further away from the truth. In particular, both the fact-checking and the media literacy intervention are *more* effective for fakes on Corona vaccines for supporters of the German AfD ("Alternative for Germany"), a far-right populist party known for spreading misinformation on Covid-19. For fakes on nutrition, where participants' beliefs are much more alike, we observe no such effect. Computing persuasion rates à la DellaVigna and Kaplan (2007) shows that this result can only

[4]While we do not find any evidence for an effect of fact-checking on the perceived credibility of non-fact-checked posts in our main experiment, a follow-up study reveals a small, but statistically significant effect. For details see Appendix B.5.

be partly explained by differences in the proportion of participants who are left to be convinced. However, we also provide evidence that AfD supporters are less certain about their prior knowledge than non-AfD supporters, so the former group may also be easier to convince. In contrast to that, we do not find any systematic effect heterogeneity in terms of education, age, social media usage, support of Corona policy measures, or prior knowledge on current events, health, and nutrition.

A plausible mechanism for our main results is that the media literacy intervention raises participants' attention and enables them to critically evaluate the Facebook postings, while fact-checking fails to markedly enhance their skills. To support the plausibility of this explanation, we show that participants who receive the media literacy intervention are more likely to actively search for further information when they respond to our questions than participants who receive the fact-checking or no intervention at all. Also, media literacy helps participants to better consider untrustworthy elements in fakes and trustworthy elements in facts. Fact-checking, in contrast, has no such effect. Moreover, we show that participants in a follow-up experiment who received a media literacy intervention spend more time reading the posts than those without any intervention, while participants who received fact-checks spend less time reading.[5]

While our main analysis illustrates the effectiveness of fact-checking and media literacy interventions in an environment where all participants see fakes and facts, it is agnostic about the extent to which the interventions are able to reverse the harm the fakes are causing. To better interpret the magnitude of our coefficients in that regard, we also compare the three main treatment groups to participants who do not see any Facebook postings at all. We find that exposure to fake news substantially impairs participants' factual knowledge, and that neither the fact-checking nor the media literacy intervention can fully offset the effect. Participants' attitudes on Corona vaccination and dietary supplements, in contrast, are hardly affected by fakes and especially the media literacy intervention can effectively repeal that impact.

To better assess the external validity of our findings, we conducted a follow-up study, applying the interventions to a new context. Specifically, we exposed participants to misinformation about environmental topics instead of Covid-19.[6] This additional experiment demonstrates that both fact-checking and media literacy interventions can be effective in different contexts and at different times. Moreover, our follow-up study indicates that the estimated effectiveness of the media liter-

[5]We did not collect the exact time spent with each post in the main experiment.

[6]Gundersen et al. (2022) and Erbaugh et al. (2024) provide reviews about online misinformation concerning the environment. For example, a lot of misinformation about climate science and climate change circulates on social media, delaying climate action (Erbaugh et al., 2024; Gundersen et al., 2022; Lewandowsky, 2021).

acy intervention is significantly greater than that of fact-checking, reinforcing our conclusion that media literacy has a broader impact than fact-checks.

As far as we know, we are the first who pursue a clean comparison of fact-checking and media literacy interventions as a means to diminish the belief in fake news, whereby we provide a valuable contribution to public and academic debates. Since public resources to combat fake news are limited, it is of utmost importance to understand when and why which remedies are most effective, so that time, money, and effort can be efficiently allocated. Pennycook and Rand (2021), for instance, stress that professional fact-checking is "simply not scalable" (p.396), as it requires substantial time and effort to examine a particular claim, and even if the claim is eventually tagged as false, the warning is likely to be missing during the peak of its spread. Similar caveats apply to the recent approach of (human) crowd sourced fact-checking (Allen et al., 2021), which is furthermore limited to the linguistic proficiency of crowd workers. AI-based procedures are more efficient than human search, but still in their infancy. Specifically, while they are able to detect potential misinformation, the fact-checking as such still requires human assessment.[7] We show that in an environment where only a small proportion of fake news can ever be fact-checked, a brief media literacy intervention is likely to be more effective than fact-checking on average. Moreover, given that displaying a small number of tips and heuristics to users of social media is cheap, scalable, and easy-to-implement, our results promote media literacy interventions as a (potentially more) powerful tool to combat fake news.

Our paper advances the surprisingly small body of research on (digital) media literacy as a means to fight fake news (A. M. Guess et al., 2020; Roozenbeek et al., 2022) and adds to a recent literature that acknowledges the limits of fact-checking (see Jerit and Zhao, 2020, for a review). E.g., Pennycook and Rand (2019) argue that many users fall for fake news because they fail to reflect; similarly, Pennycook et al. (2021) and Pennycook et al. (2020) show that users share false claims partly because they do not think sufficiently about whether or not the content is accurate. Guriev et al. (2023) show that an attention nudge (which could be thought of as a minimalistic media literacy enhancement) can reduce sharing of fake content on X to a greater extent than fact-checks. Consistent with what we find, such results advocate media literacy interventions that help users to critically evaluate social media postings as a promising avenue, while assorting fact-checking – which fails to markedly enhance users' skills – as less effective.

The remainder of the paper is organized as follows. Section 3.2 reviews the related

[7]See EU Horizon, URL: `https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news` (viewed August 2023).

literature on social media, misinformation, and education interventions. Section 3.3 illustrates the experimental setup and implementation; moreover, we describe our empirical strategy. Section 3.4 presents our main results, where we compare the effectiveness of fact-checking and media literacy on the credibility of and factual knowledge on fakes, as well as on participants' attitudes. In Section 3.5, we show that an increase in attention and the ability to critically evaluate social media postings on behalf of the media literacy intervention is a plausible mechanism for our results. Section 3.6 presents further results and robustness checks, Section 3.7 concludes. We describe our follow-up experiment which includes stimuli on environmental topics in Appendix B.5.

3.2 Related literature

Social media and UGC Our paper is related to two strands of literature. First, it adds to the vibrant and interdisciplinary research on social media and user-generated content (reviewed by Luca, 2015; Zhuravskaya et al., 2020), where it is particularly close to analyses of fake news. This subfield can be further divided into studies on the emergence and spread of fake news (e.g., Allcott and Gentzkow, 2017; Bursztyn et al., 2023; Grinberg et al., 2019; A. Guess et al., 2019; A. Guess et al., 2018; D. M. Lazer et al., 2018; Vosoughi et al., 2018), and inquiries of potential remedies (reviewed by Allen et al., 2021; Jerit and Zhao, 2020; Lewandowsky et al., 2012).[8] The latter literature focuses on corrective interventions like fact-checking: While Bode and Vraga (2015), Vraga and Bode (2017), and Henry et al. (2022), among others, support its effectiveness, other papers find no or even “backfire” effects (e.g., Nyhan and Reifler, 2010; Nyhan and Reifler, 2015), or they document mixed results, whereby fact-checking improves users' factual knowledge, but struggles to change more deep-rooted perceptions and attitudes (Barrera et al., 2020; Nyhan et al., 2020). Indeed, the efficacy of fact-checking varies significantly depending on the outcome being considered. The consensus in the literature suggests that while fact-checking has limited effect on correcting fundamental beliefs, it is effective in influencing actions, such as reducing the sharing rate of false news. Ex post fact-checking, in particular, may be ineffective, as it is challenging to correct beliefs after exposure to false or misleading information (Barrera et al., 2020; Nyhan et al., 2020; Swire et al., 2017). Conversely, Henry et al. (2022) demonstrate that both imposed and voluntary fact-checking reduce the sharing of false statements by over 25%. Recent work by

[8]Because of the topics of the statements we use as experimental stimuli, our main experiment is particularly related to the literature on misinformation about health (see Swire-Thompson and D. Lazer (2020) for a review), and our follow-up study to the literature on misinformation about the environment (see Gundersen et al. (2022) and Erbaugh et al. (2024) for reviews).

Drolsbach et al. (2024) extends this discussion by examining X’s community-based fact-checking system “Community Notes”. Their findings highlight that explanatory context provided by community notes not only fosters trust in fact-checks across political divides, but also enhances users’ ability to discern misleading content.

The literature on fact-checking varies in terms of whether corrections are provided before or after exposure to false information. Both approaches appear to be somewhat effective, but there is no clear consensus on which is superior (Swire-Thompson et al., 2021). For instance, Lewandowsky et al. (2012), U. K. Ecker et al. (2022), and Lewandowsky and Van Der Linden (2021) advocate for *pre-bunking*, where corrective information is presented before exposure to fakes. They argue this method more effectively prevents the “continued influence effect” (Lewandowsky et al., 2012), where false information lingers in memory once learned. However, other studies find that the order of presenting fakes and corrections does not significantly impact effectiveness (Swire-Thompson et al., 2021), or that corrections shown after the fakes have a greater effect (Brashier et al., 2021; Dai et al., 2021). In our study, we align with Facebook’s current practice by presenting corrections before showing the fake information to participants.[9]

A novel line of research examines the efficacy of removing misleading content or platform users altogether. E.g., Broniatowski et al. (2023) and Gu et al. (2022) evaluate the impact of Facebook’s strict misinformation policy in March 2019 on user endorsements of vaccine content. Similarly, Mitts et al. (2022) and Ali et al. (2021) evaluate the effectiveness of banning prominent groups and figures who spread false or misleading content on Facebook and X. These studies agree that removing content or users is not very effective, as it does not reduce overall engagement with misleading content, and arguably increases the production of fake news elsewhere. In contrast, Ershov and Morales (2024) evaluate a user-centric intervention where Twitter introduced friction to content sharing during the 2020 U.S. presidential election.

Studies on alternative ways to combat fake news are relatively rare. One notable exception is A. M. Guess et al. (2020), who assess the effectiveness of Facebook’s “Tips to Spot False News” on discernment between mainstream and false news headlines both among a nationally representative sample in the US and a highly educated online sample in India. Relatedly, Roozenbeek et al. (2022) use five short videos that inoculate people against manipulation techniques commonly used in misinformation and find that they improve manipulation technique recognition, boost confidence in spotting these techniques, increase users’ ability for truth discernment as well as

[9] In a follow-up experiment, we randomize the order in which participants view the fake and the corresponding fact-check. For more details see Appendix B.5.3.4.

the quality of their sharing decisions. Bak-Coleman et al. (2022) derive a generative model of fake news engagement and use Bayesian simulation techniques to show that fact-checking, removal of misleading content and users, and nudges towards accuracy are efficient in combination, but unlikely to work in isolation. Guriev et al. (2023) compare the effect of an attention nudge (which could be thought of as a minimalistic media literacy enhancement) to fact-checking and show that it is more effective than fact-checking and other policies in reducing the sharing rates of fake posts on X.

We contribute to this literature in several ways. First, we pursue a clean comparison of the effect of fact-checking and media literacy interventions on the belief of consumers, which has not been done so far. In particular, our experimental setup allows us to study the immediate and short-term effects of fact-checking and media literacy interventions in one and the same environment, whereby we can observe when and why which remedy is most effective. In addition, we provide evidence for potential mechanisms behind our results, shifting the research focus from asking *whether* fact-checking and media literacy interventions are effective tools to fight fake news to studying *how* they work and *in which case* they fail or succeed.

Very closely related to our study are Barrera et al. (2020) and A. M. Guess et al. (2020). Barrera et al. (2020) use a randomized online experiment to expose voters to fakes, facts, and fact-checks on immigration in France. Participants are then asked about their posterior beliefs on topics related to immigration, their opinions on immigration policy, as well as their voting intentions. Similar to what we find, Barrera et al. (2020) demonstrate that fake news are highly persuasive, and while fact-checking enhances factual knowledge, it fails to offset the fakes' effect on voting intentions.[10] A. M. Guess et al. (2020) examine the impact of a digital media literacy intervention on the perceived accuracy of false and correct news headlines and show that participants' ability for truth discernment increases.

Our results largely confirm these findings, but we extend the preceding analyses in several ways. First, we explore the immediate *and* short-term effects of fact-checking *and* media literacy interventions on fakes *and* facts within one experiment, which allows us to directly compare these remedies and draw a sophisticated picture of how and when which type of intervention works. Likewise, we consider a broad range of coherent outcomes – credibility, factual knowledge, and attitudes – and complement our analysis with a thorough examination of potential mechanisms. Finally, we use postings from social media that actually exist and whose content is not necessarily politically loaded, demonstrating that our results hold beyond the partisan context.

We also contribute to a growing body of research arguing that users fall for fake news because they fail to pay sufficient attention (e.g., Loewenstein and Wojtowicz,

[10] Similar results are presented by Nyhan et al. (2020).

2023; Pennycook and Rand, 2019). Pennycook et al. (2021) and Pennycook et al. (2020), for instance, show that users frequently share misinformation because they do not focus on accuracy; politically motivated reasoning, in contrast, seems to play a minor role. Our results are in line with such findings, because we demonstrate that media literacy interventions – which raise users' attention and help them to actively distinguish between fakes and facts – are on average more effective than just passively receiving fact-checks. Moreover, in contrast to previous findings on motivated reasoning (e.g., Jerit and Zhao, 2020; Lewandowsky et al., 2012), we find that our interventions are more effective for supporters of a far-right political party, who are initially much more likely to oppose Corona vaccination. This, too, is consistent with the above line of thought, whereby it is often a lack of attention rather than partisanship why people fall for fake news.

Education interventions Second, our paper is related to the literature on education interventions, most of which focuses on financial literacy. In accordance with our results on the media literacy intervention, a meta-analysis of 76 randomized experiments by Kaiser et al. (2022) reveals that financial education interventions have, on average, positive causal treatment effects on financial knowledge and behavior. The treatment effects are economically meaningful in size and comparable to those realized by education interventions in math and reading (e.g., Cheung and Slavin, 2016; Fryer Jr, 2017; Hill et al., 2008), health (e.g., Noar et al., 2007; Rooney and Murray, 1996), and energy saving behavior (e.g., Karlin et al., 2015). In particular, Kaiser et al. (2022) report that the average effect size of financial education interventions corresponds to about 0.123 standard deviation units. This is only about half as much as what we see for the impact of our media literacy intervention on the credibility of fakes, and roughly as much as what we observe for the impact on factual knowledge on average. However, Kaiser et al. (2022) stress that the effectiveness of education interventions diminishes over time. Given that the average effect size from their meta-analysis also includes interventions that were evaluated after several weeks or months, it is not surprising that the effect sizes from our media literacy intervention are comparably large.

A plausible explanation for the larger effectiveness of the media literacy relative to the fact-checking intervention is that the former raises participants' attention and enables them to critically evaluate the postings' accuracy, whereas the latter only helps them to update beliefs about one specific fact and fails to markedly enhance their more general skills. This matches the findings by Kaiser and Menkhoff (2022), who study different types of interventions offered to small-scale retailers in Uganda and show that "active learning" has a positive effect on savings and investment outcomes while traditional lecturing is ineffective. Comparable results on the advantages of active as opposed to passive learning have been documented in other domains, including science, technology, engineering, and maths (e.g., Deslauriers et al., 2011; Freeman et al., 2014; Ruiz-Primo et al., 2011). Relatedly, Drexler et al. (2014) show that a heuristics-based approach relying on "rules-of-thumb-training" – such as our ten tips to spot false news – generates larger behavioral impacts than the teaching of full curricula.

3.3 Experimental design

3.3.1 Survey flow

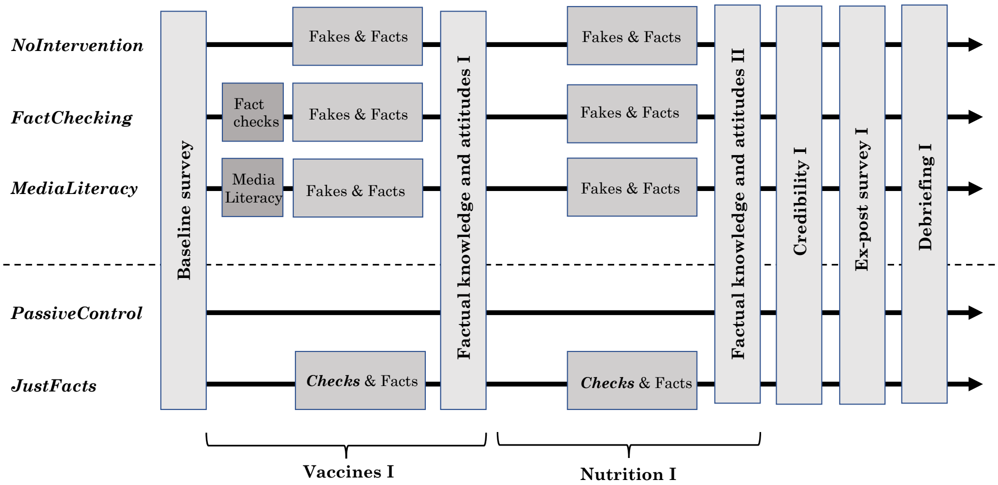
We start by randomizing the participants of our online survey experiment into one out of five groups of approximately equal size: (i) NoIntervention, (ii) FactChecking, (iii) MediaLiteracy, (iv) JustFacts, and (v) PassiveControl. To study both the immediate and short-term effects of our interventions, we conduct two waves of the experiment, where we re-invite the same participants one week after they completed Wave I and allocate them to the same treatment group as before. Figure 3.1 gives an overview of our survey flow, further details are discussed below.[11]

3.3.1.1 Wave I

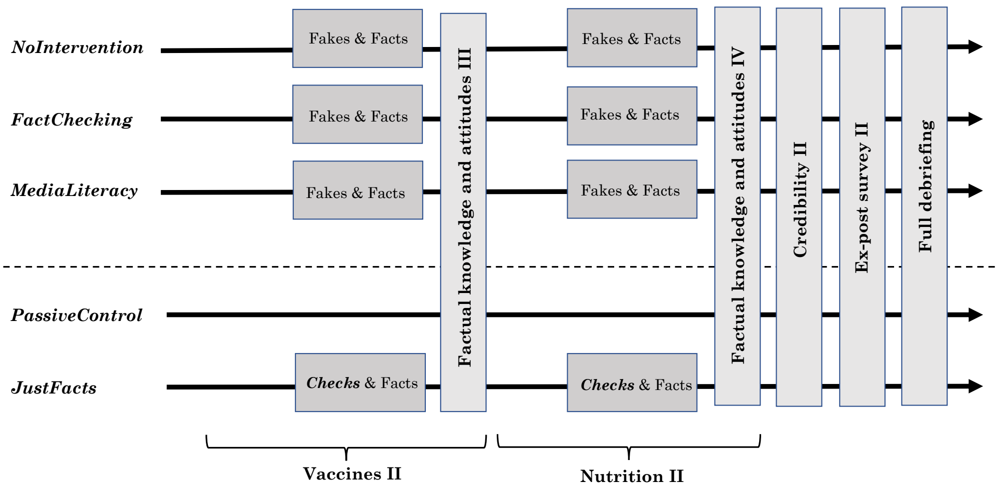
Baseline survey All participants start with a baseline survey on standard demographics such as age, gender, family status, household income, education, profession, and personality traits ("big five"). In addition, we inquire participants' prior knowledge on current events, health, and nutrition.[12] To avoid priming effects on subsequent questions, we do not include questions about the figures of interest after the interventions, but ask (i) how many days Joe Biden has been President of the United States, (ii) when to see a doctor in case of high temperature, and (iii) how many servings of fruit and vegetables are officially recommended per day. To measure the strength of participants' prior beliefs, we also ask how certain they are about the accuracy of their responses on a 5-point Likert scale ranging from *Very uncertain* to *Very certain*. To maintain participants' engagement, we incorporated two attention checks: one placed in the middle of the baseline survey and another at the end.

[11]We preregistered our two interventions and five treatment groups as described above. We sketched the survey flow but did not preregister it in full detail. We also did not preregister which fakes, facts, and fact-checks we would display exactly. Please see Appendix B.2 for details.

[12]We preregistered the full set of control variables. Please see Appendix B.2 for details.



(a) Overview Wave I



(b) Overview Wave II

Figure 3.1: Survey flow

Vaccines Participants in the NoIntervention, the FactChecking, and the MediaLiteracy group are shown two pieces of "fake news" ("fakes" henceforth) and two facts on Corona vaccines in randomized order.[13] The fakes and facts were manually collected from Facebook, i.e., we use screenshots of Facebook postings that actually exist. All fakes and facts were collected between March and May 2021.

We used two search strategies to collect the postings. First, we used Facebook's search function with German keywords equivalent to "covid", "corona", "vaccine" and "side effect". We then screened through all publicly available search results, including postings from individual profile pages and Facebook groups. Second, we started from the fake news repositories of renowned German fact-checking initiatives like *AFP* and *Correctiv.org*. Specifically, we skimmed through all registered fake news and checked which of them stem from Facebook and deal with the topic of interest. The collected postings from either search strategy had to meet the following criteria: They were only included in the experiment if we could find appropriate fact-checks debunking the false information, they had to be either about Covid-19 vaccines or dietary supplements and they had to be posted not too long ago. In addition, all fakes and facts had to contain a concrete numerical value (e.g., "50 people died after vaccination in a Sana clinic") that we could later on ask for. Using this catalogue of criteria led to the pool of posts we use. See Appendix B.1.3 for all fakes and facts that are part of the experiment.[14]

Participants in the NoIntervention group do not receive further information. Participants in the FactChecking group, in contrast, receive additional fact-checks that explicitly debunk the false information (e.g., an official statement that the story about 50 deaths after vaccination in a Sana clinic is false). All fact-checks stem from sources that are commonly perceived as trustworthy (e.g., *Correctiv.org*). The fact-checks are shown *prior* to the fakes that they correct. We thereby follow the current procedure on Facebook, where false or misleading information – if detected – is overlain with a warning message that redirects the user to a fact-check; the original post can only be seen after the user closes the warning. Moreover, given that the timing of the intervention is known to be relevant (Brashier et al., 2021; Lewandowsky et al., 2012), displaying the fact-check prior to the respective fake makes the fact-checking better comparable to the media literacy intervention. See Appendix B.1.3 for all fact-checks that we use.[15]

[13]We find no evidence for order effects: Our main results are unaffected when we control for the order of the fakes and facts. Moreover, we find no differences between participants who first saw a fake and participants who first saw a fact.

[14]To strengthen the external validity of our findings, a follow-up experiment exposes a new sample of participants to fakes and facts on environmental topics (see Appendix B.5.3.2).

[15]We did not preregister whether we would display the fact-checks before or after showing the respective fake (see Appendix B for details). In our main experiment, we show the fact-check *before*

Participants in the MediaLiteracy group receive Facebook's official "Tips to spot false news" *before* they are exposed to fakes and facts about Corona vaccines.[16] These tips actually exist on the platform and comprise ten short pieces of advice, including "Be skeptical of catchy headlines", "Look closely at the link", and "Investigate the source"; Appendix B.1.1 shows the full list. We inform our participants that these tips have been developed by Facebook itself. We display one tip per page and ask the participants to read them carefully before they proceed to the Facebook postings on Corona vaccines.

"Media literacy" is commonly understood as the ability to access, analyse, understand, and reflect on messages conveyed through mass media.[17] Our brief intervention, which offers ten tips to spot false news, cannot fully equip an individual with all the skills required for complete media literacy. However, these tips provide practical heuristics that enhance participants' awareness and aid in their critical evaluation of the fakes and facts that we present. Thus, while our "media literacy intervention" cannot fully develop media literacy, it reinforces existing skills and encourages participants to apply them effectively.

Note that participants in the NoIntervention group do not receive any placebo treatment such as short pieces of text with unrelated information (e.g., on financial education). First, going through the fact-checking or the media literacy intervention requires a relatively short amount of time, so we do not expect any differences in performance between the treatment groups owing to differences in fatigue. Second, we consider the scenario where participants read fakes and facts and nothing else as the most relevant and externally valid benchmark. In particular, we wish to examine if the fact-checking or the media literacy intervention help participants to better distinguish between fakes and facts relative to a situation where they receive no treatment at all, as would often be the case on social media platforms.

In contrast to the other groups, participants in the JustFacts and in the PassiveControl group are *not* exposed to fakes. While the PassiveControl group does not see any postings at all, participants in the JustFacts group receive the same two facts and fact-checks (without the corresponding fakes) as participants in the FactChecking group.[18] We can thereby infer our participants' average prior beliefs and attitudes from responses by the PassiveControl, and the impact of stand-alone fact-checks from the JustFacts group.

the respective fake. Appendix B.5.3.4 shows results for a follow-up experiment where we randomize the order of fact-checks and fakes.

[16] These tips have been developed in cooperation with several professional fact-checking initiatives. See `https://www.facebook.com/help/188118808357379` (viewed December 2021).

[17] See, e.g., the European Commission: `https://digital-strategy.ec.europa.eu/en/policies/media-literacy` (Viewed: June 2024).

[18] The fact-checks are self-contained and can stand on their own.

After exposure to the Facebook postings, we ask all participants four factual questions that are tailored to the fakes and facts just shown (see Appendix B.1.3 for all fakes and facts):

1. How many employees of Sana Kliniken died after being vaccinated against Corona in March 2021?

2. How much radioactive radiation does a dose of Corona vaccine emit (in microsievert)?

3. How effective is the Biontech/Pfizer vaccine in adolescents (in percentage)?

4. How many doses of the Russian Sputnik V vaccine has Bavaria ordered independently of the EU?

Each question asks for a specific number, and participants must give their answer through an input box, i.e., we do not provide a list of pre-defined options. To secure high quality answers, we use a bonus payment scheme that rewards participants whose answers are close to the true value.[19]

Next, we inquire all participants' willingness to get vaccinated against Covid-19. We start by asking each participant if he or she was already fully vaccinated by the time of the experiment. Based on their response, we partition the group into those who are fully vaccinated and those who are not. Then, we ask the former group about their willingness to get a booster injection as soon as it is officially recommended, and the latter about their willingness to get vaccinated against Covid-19 in general. Answers could be given on a 5-point Likert scale ranging from *Very likely* to *Very unlikely*. To avoid experimenter demand effects, we do not incentivize this question with a potential bonus payment (see Section 3.6.6.3 for further discussion). To maintain participants' engagement, we placed an attention check before proceeding to the second part of the survey.[20]

Nutrition The second part of Wave I is analogous to part one, except that we switch from Corona vaccines to nutritional topics, and that there are no further interventions (i.e., the setup is identical for participants in the NoIntervention,

[19]More specifically, we use a quadratic scoring rule, whereby answers close to the true value increase participants' chance to receive a bonus payment of 20 EUR.

[20]We ask participants about their vaccination status at the latest possible point in the experiment to avoid priming effects. However, it is possible that the experimental stimuli could increase participants' self-reported vaccination rates. If this was true, vaccination status cannot be used as a control variable. We are not overly concerned with this possibility since in this case we would expect the vaccination rate in the PassiveControl group to be as low as in the NoIntervention group. Yet, as shown in Table B1 in Appendix B.4, this is not the case. To ensure full transparency, we present the results both with and without using vaccination status as a control variable.

the FactChecking, and the MediaLiteracy group). For this part, we used German keywords equivalent to "protein", "vitamin", "nutrition", "diet" and "dietary supplement" to find the experimental stimuli. The main idea is to explore if the fact-checking and the media literacy interventions stay effective in a different context that is health-related, too, but unlikely to be influenced by politically motivated reasoning.[21] As before, participants in the NoIntervention, the FactChecking, and the MediaLiteracy group are shown two fakes and two facts on nutritional topics in randomized order; these fakes and facts have to fulfill the same requirements as above. All fakes on nutrition promote the intake of needless dietary supplements such as extra protein or Vitamin C. Participants in the PassiveControl and the JustFacts group are not exposed to fakes on nutrition, but the latter receive two facts and two fact-checks.[22]

Analogous to part one of the survey, we proceed with a quiz that comprises four factual questions tailored to the fakes and facts that have just been shown:

1. How much protein should a person consume daily (in grams per kilogram bodyweight)?
2. What percentage of adults suffer from a lack of Vitamin C?
3. How many grams of microplastics does a person consume on average per week?
4. How many percent of German adolescents exercise less than recommended by the World Health Organization?

Again, each of those questions asks for a specific number, answers must be given through an input box, and we remind our participants of the potential bonus payment to incentivize high quality answers.

Finally, we inquire all participants' willingness to consume dietary supplements, where they can respond on a 5-point Likert scale ranging from *Very likely* to *Very unlikely*. To maintain participants' engagement, we placed an attention check before proceeding to the final part of the survey.

Credibility Next, we inquire the perceived credibility of all fakes, facts, and fact-checks. To this end, we display all postings again and let participants rate their

[21] Though not as topical as Corona vaccines, the consumption of (needless) dietary supplements is an important concern. Recent surveys indicate that nearly 50% of all German adults have purchased dietary supplements within the last six months, but almost a third of them feels ill-informed about potential health risks that go along with their consumption (Verbraucherzentrale, 2022). Moreover, the consumption of dietary supplements does typically not go along with improved public health (Radimer et al., 2004) – quite the contrary – as dietary supplements are often either ineffective (DGE, 2012) or even harmful (Chiou et al., 2011).

[22] Note that the FactChecking group does *not* receive these fact-checks.

credibility on a 5-point Likert scale ranging from *Very credible* to *Very incredible*. Participants are only asked about postings that they saw during the experiment, i.e., the FactChecking group is asked about fakes, facts, and fact-checks, the MediaLiteracy and the NoIntervention groups are asked about fakes and facts, the JustFacts group is asked about facts and fact-checks, and the PassiveControl group is not asked at all. We deliberately inquire the credibility of fakes, facts, and fact-checks at this late stage of the experiment to avoid priming effects on the preceding questions. Moreover, to avoid experimenter demand effects, the credibility questions are not incentivized with a potential bonus payment, and we explicitly state that there is "no correct answer" and that we are "interested in [the participants'] personal opinion".

Ex-post survey The ex-post survey helps us better understand potential mechanisms and gather information we intentionally omitted earlier to avoid priming effects. We proceed in three steps:

(Un-)trustworthy elements First, we investigate whether enhanced awareness and critical thinking are plausible mechanisms driving differences between treatment groups. We assess whether participants improve at considering trustworthy elements in facts (e.g., a reputable news source) and untrustworthy elements in fakes (e.g., spelling mistakes), but not vice versa, ensuring that they make conscious and correct choices. To that end, we display all fakes, facts, and fact-checks again and let participants indicate which elements of each posting they perceive as particularly trustworthy or untrustworthy.

To obtain comparable responses, we pre-define five categories of elements in the postings: (i) *format and spelling*, (ii) *information as such*, (iii) *images*, (iv) *source and URL*, and (v) *verified account*. The category *format and spelling* includes all grammar, spelling, and punctuation mistakes, as well as overly many question and exclamation marks, or uncommon fonts and colors. Elements from the category *information as such* comprise numbers and dates in a posting as well as all further "factual" information (e.g., "we are all suffering from a lack of vitamins"). *Images* refers to all photographs, graphs, and other illustrations in a posting. The category *source and URL* comprises both the author of the posting, including his or her profile name and picture, and the claimed source of the information presented, ranging from "my friend said" to "a study from MIT". Finally, the *verified account* category refers to Facebook's blue check mark. Note that some of the postings do not exhibit elements from each category; e.g., some of them feature no image, no verified account label, or no spelling mistakes. Hence, the number of elements that participants can consider as trustworthy or untrustworthy is different for each posting. See Appendix B.1.4 for some illustrative examples.

Our pre-defined elements are not immediately visible for the participants. Only as soon as they hover their cursor over a pre-defined element – say, a spelling mistake – the respective element gets shaded and participants can either single-click (indicate untrustworthy) or double-click (indicate trustworthy) on it. Participants can consider none, one, or several elements per posting as either trustworthy or untrustworthy. Participants must complete a mandatory tutorial to get familiar with the procedure before they can proceed. As with the credibility questions, participants are only asked about postings that they saw during the experiment (e.g., only participants in the FactChecking group are asked to consider (un-)trustworthy elements in fact-checks).

Further questions Second, we ask further questions on participants' political party preferences, social media usage, if they got vaccinated against any type of disease during the past ten years, and if they agree with the current Corona regulations. Moreover, we ask all participants if they searched for further information online.

List experiments Third, we conduct two list experiments à la Blair and Imai (2012) to check whether our main results on attitudes are driven by a "Bradley effect" (Hopkins, 2009), whereby participants conceal socially undesirable opinions and attitudes (see Section 3.6.6.3 for further details).[23]

Debriefing At the end of Wave I, we debrief all participants by displaying the correct answers to the factual questions on Corona vaccines and nutrition.

3.3.1.2 Wave II

To study the effectiveness of our fact-checking and media literacy interventions in the short-run, we re-invite all participants after about one week to Wave II of the experiment. Wave II replicates the steps from Wave I, except that there are no further interventions (i.e., the setup is identical for the No Intervention, the Fact-checking, and the Media Literacy groups), no baseline survey, and that we use a different set of fakes, facts, and fact-checks.[24]

[23] Our list experiments were not pre-registered.

[24] The factual questions on Covid-19 vaccines are: (i) How many European countries have completely taken the AstraZeneca Corona vaccine off the market?, (ii) In which year was the vaccine shown here manufactured?, (iii) How many doctors missed a scheduled vaccination appointment in Saarland in February?, and (iv) How many compensation claims for complications from a Corona vaccination were submitted to the state of NRW by the end of June?. The factual questions on nutrition are: (i) What percentage of Germans have a Vitamin B12 deficiency?, (ii) How much

We conclude the survey with a full debriefing of all participants. To this end, we display the correct answers to all factual questions, show all fact-checks to all fakes that were used during the survey, and provide links to trustworthy websites on Corona vaccines and nutrition, where the participants can get further information on these topics if they wish.[25]

3.3.2 Implementation

The experiment was programmed with the survey software *Qualtrics* and conducted in cooperation with *respondi*, a major commercial panel provider.[26] We used e-mails to invite around 3,000 German participants to Wave I of the survey, i.e., around 600 participants per group.[27] Participants had to be between 18 and 59 years old; conditional on that requirement, the sample is representative for the German population in terms of gender, age, and state of residency.[28] Participants could use their smartphones, tablets, or desktop PCs to answer our questions. Those who completed the survey received the usual payment by *respondi* plus the potential bonus payment.[29]

We conducted Wave I of the experiment between September 9th and September 29th in 2021, and Wave II between September 26th and October 27th. Participants received a re-invitation about one week after they completed Wave I. The minimum interval of actual participation in the two waves is equal to eight days, though, the median interval is equal to 15, and the mean interval equal to 17.6 days (see Section 3.6.6.2 for further discussion). The response rate in Wave II is equal to 83% – which is roughly equal to *respondi*'s average – and we find no evidence for differential attrition.[30] At the time the experiment took place, all German adults had had the

Vitamin E should an adult in Germany consume per day on average (in milligrams)?, (iii) What percentage of sugar does a German child consume on average too much?, and (iv) How many kilocalories are in 100 grams of chia seeds?.

[25]Specifically, we suggest to visit the websites of the *Robert Koch Institut (RKI)* (URL: `https://www.rki.de/DE/Home/homepage_node.html`) and the *National Ministry of Health* (URL: `https://www.bundesgesundheitsministerium.de/`) further information on Corona vaccines, and to visit the website of the *German Agency for Nutrition* (URL: `https://www.dge.de/`) for further information on nutrition.

[26]Cooperating with professional panel providers such as *respondi* has become standard in economic research; see, e.g., Stantcheva (2021) and Alesina et al. (2023) for examples and `https://www.bilendi.com/` for further details on *respondi*.

[27]The sample size is comparable to related studies, e.g. Barrera et al. (2020) and Henry et al. (2022).

[28]Although our participants are likely to encounter misinformation on Corona vaccines frequently in their every day lives, we exclude participants aged 60+ as an especially vulnerable group from our experiment.

[29]Our sample size and expected attrition rate were preregistered; see Appendix B.2 for details.

[30]Response rates per group: NoIntervention 82.36%, FactChecking 84.51%, MediaLiteracy 83.36%, JustFacts 78.90%, PassiveControl 84.62%.

opportunity to get fully vaccinated (two injections), and policy makers and health experts were discussing whether and when a third injection would make sense.

3.3.3 Balance check

Table B1 in Appendix B.4 displays the means and standard deviations of all control variables for each treatment group. Since we use the NoIntervention group as baseline in the subsequent analyses, we also conduct t-tests on the difference in means between the NoIntervention and each of the other treatment groups, respectively.

We find that our sample is strongly balanced with respect to age, gender, family status, state of residence, consumption of dietary supplements, and prior knowledge on current events, health, and nutrition, but there are small differences between some of the treatment groups for household income, education, party preferences, and Corona vaccination status. To take these imbalances into account, we include the full set of pre-registered control variables into each of our regression analyses.[31] Since we did not pre-register participants' Corona vaccination status as a control, we only include it as a robustness check – with one exception (see Section 3.4.1.3) our results are unaffected.[32]

3.3.4 Variables

Next, we aggregate our participants' responses to the various fakes and facts and convert them into measures suitable for regression analyses. We also standardize responses to the prior knowledge questions and generate an indicator for participants' uncertainty about them. Table B2 provides summary statistics of all dependent variables that we use in the analysis.[33]

Credibility We start by computing each participant's mean response to the credibility questions on Corona vaccine fakes, Corona vaccine facts, nutrition fakes, and nutrition facts for each of the two waves, respectively (i.e., we compute eight mean responses per participant). Then, we define a dummy variable equal to one if the mean response indicates that the participant perceives the fakes or facts on average as *Very credible*, *Credible* or *Indecisive*.[34] This aggregation level allows us to

[31] The preregistered control variables are age, gender, family status, state of residence, personality traits ("big 5"), household income, education, party preferences, and prior knowledge on current events, health, and nutrition; see also Appendix B.2.

[32] For a detailed discussion of heterogeneity in terms of vaccination status see section 3.6.6.1.

[33] We preregistered our main outcome variables as such, but not that we would measure them on a 5-point Likert scale and transform them to binary outcomes or standardized values (see Appendix B for details). However, our results are robust to alternative specifications of the outcome variables.

[34] Our results are robust to alternative cutoffs and to using the average response on the 5-point Likert scale prior to its transformation to a binary measure. Even with an extreme cutoff, where

examine the treatment effect on fakes and facts separate from each other, whereby we can show that our interventions have no detrimental effect on facts.

Factual knowledge Next, we standardize participants' responses to the factual knowledge questions. To this end, we first compute the absolute distance between each response and the correct answer. E.g., the correct answer to "How many people died after vaccination in a Sana clinic?" is equal to zero; if the participant's response is "50", the absolute distance between response and correct answer is equal to 50. To avoid distortion through outliers, we winsorize all distances to each question at their 95^{th} percentile.[35] Then, based on the entire sample, we standardize all winsorized distances to have a mean of zero and a standard deviation of one, which allows us to compare responses across questions. Finally, we aggregate participants' responses by computing their mean standardized distance to the correct answer to the factual questions on Corona vaccine fakes, Corona vaccine facts, nutrition fakes, and nutrition facts for each of the two waves, respectively (i.e., we compute eight mean responses per participant again).

Attitudes To capture participants' attitudes, we define a dummy that is equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19 in each of the two waves, respectively. Analogously, we define a dummy equal to one if he or she states to be *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the near future.[36]

(Un-)trustworthy elements in fakes and facts To measure how much attention participants pay to the content of the fakes and facts and how critically they evaluate them, we count how many elements (in absolute terms) they consider as trustworthy or untrustworthy in each posting. Then, for each participant, we compute the average number of (un-)trustworthy elements that he or she considered for Corona vaccine fakes, Corona vaccine facts, nutrition fakes, and nutrition facts for each of the two waves, respectively.

Prior knowledge Analogous to factual knowledge, we make participants' responses to each of the three prior knowledge questions better comparable by computing the

the indicator variable is set to 1 only if participants rated fakes as *very incredible*, the findings remain consistent (taking into account the reverse coding).

[35]Note that we pre-registered our intention to winsorize each participant's responses at their 95^{th} percentile. We obtain similar results when we drop outliers beyond the 95^{th} percentile, winsorize responses at their 99^{th} percentile, or do not winsorize at all. The distribution is based on the entire sample.

[36]Our results are robust to alternative cutoffs and to using the average response on the 5-point Likert scale prior to its transformation to a binary measure.

standardized distance between a participant's response and the correct answer. In addition, we compute an indicator that is equal to one if participant i is on average *Very uncertain*, *Uncertain*, or *Somewhat certain* about his or her prior knowledge.

3.3.5 ITT analysis

In our baseline analysis, we estimate the Intention to Treat Effects (ITT) of our interventions by using OLS to estimate the regression equation

$$y_{iw} = \beta_0 + \beta_1 TG_i + \beta_2 X_i + \varepsilon_{iw}, \tag{3.1}$$

where y_{iw} corresponds to an outcome of participant i in survey wave w as described above, TG_i denotes participant i's treatment group, and X_i is a vector of pre-registered control variables including age, gender, party preferences, religion, education, family status, household income, personality traits, state of residence, and prior knowledge on current events, health, and nutrition. The baseline category in TG_i is the NoIntervention group, i.e., we compare participants who receive fakes and facts without further intervention to participants in each of the other treatment groups.[37]

We refer to these estimates as ITT, because participants from the FactChecking and the MediaLiteracy group could theoretically skip the intervention by just quickly clicking through the survey. We show in Section 3.6.6.4 that our findings are robust to an IV-specification.

3.4 Results

The main purpose of our paper is to study whether and when fact-checking and media literacy interventions are able to debunk fake news that circulate on social media. To this end, we focus on comparing the NoIntervention to the FactChecking and the MediaLiteracy group here and defer supporting analyses of the JustFacts and the PassiveControl group to Section 3.6.

We consider three types of outcomes: the credibility of fakes and facts, factual knowledge on the topics the fakes and facts are dealing with, and attitudes towards Corona vaccination and the intake of dietary supplements. The idea is to examine a coherent cognitive chain: Do the interventions reduce the credibility of fakes (but not of facts)? If yes, does that translate into better factual knowledge on the topics

[37]We preregistered our baseline OLS regression estimation as described above. See Appendix B.2 for details.

the fakes are dealing with? If yes, does that affect participants’ attitudes? Table 3.1 displays all mean outcomes per treatment group.

3.4.1 ITT analysis

We start by examining the ITT of our interventions (Section 3.6.6.4 presents an IV analysis). In particular, we demonstrate that the effectiveness of fact-checking tends to be limited to the fakes that are corrected, while the media literacy intervention helps to distinguish between fakes and facts more generally. Figures B8 to B10 in Appendix B.3 illustrate the results for each outcome, treatment, and wave of the survey; further details are presented below.

Table 3.1: Mean values of all outcomes per treatment group

Panel A: Wave I	**Credibility**		**Knowledge**		**Truth discernment**				**Attitudes**	
					Credibility		Knowledge			
	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No Intervention	0.299	0.901	0.398	0.132	0.584	0.922	0.152	0.016	0.783	0.557
	(0.018)	(0.012)	(0.034)	(0.025)	(0.014)	(0.008)	(0.024)	(0.020)	(0.017)	(0.020)
Fact-checking	0.219	0.886	0.076	0.051	0.538	0.913	0.021	-0.032	0.830	0.555
	(0.017)	(0.013)	(0.033)	(0.025)	(0.014)	(0.008)	(0.023)	(0.019)	(0.015)	(0.020)
Media Literacy	0.195	0.842	0.179	0.053	0.529	0.896	0.047	-0.013	0.836	0.515
	(0.016)	(0.015)	(0.034)	(0.026)	(0.014)	(0.009)	(0.023)	(0.020)	(0.015)	(0.020)
Panel B: Wave II	**Credibility**		**Knowledge**		**Truth discernment**				**Attitudes**	
					Credibility		Knowledge			
	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition	Covid	Nutrition
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No Intervention	0.403	0.831	0.334	0.143	0.635	0.896	0.077	-0.031	0.741	0.554
	(0.022)	(0.017)	(0.037)	(0.026)	(0.015)	(0.010)	(0.025)	(0.021)	(0.019)	(0.022)
Fact-checking	0.394	0.806	0.286	0.164	0.625	0.872	0.081	-0.022	0.797	0.565
	(0.022)	(0.018)	(0.039)	(0.028)	(0.015)	(0.010)	(0.025)	(0.022)	(0.018)	(0.022)
Media Literacy	0.350	0.759	0.220	0.117	0.615	0.861	0.036	-0.047	0.804	0.532
	(0.021)	(0.019)	(0.037)	(0.028)	(0.015)	(0.011)	(0.025)	(0.022)	(0.018)	(0.022)

Notes: Table 3.1 shows the mean outcome values for each of our treatment groups, respectively. Standard errors in parentheses.

3.4.1.1 Credibility

Table 3.2 presents the regression results for the perceived credibility of fakes. Panel A shows the estimates from comparing the FactChecking, and Panel B from

Table 3.2: Credibility of fakes

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.080	-0.071	-0.015	-0.010	-0.009	0.016	-0.026	-0.018
	[0.025]	[0.024]	[0.018]	[0.017]	[0.031]	[0.030]	[0.024]	[0.024]
p-value	(0.001)	(0.004)	(0.385)	(0.553)	(0.769)	(0.582)	(0.282)	(0.460)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	-0.104	-0.105	-0.060	-0.061	-0.052	-0.042	-0.072	-0.068
	[0.024]	[0.024]	[0.019]	[0.019]	[0.030]	[0.029]	[0.025]	[0.025]
p-value	(0.000)	(0.000)	(0.002)	(0.001)	(0.084)	(0.143)	(0.004)	(0.006)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020
Baseline: No Intervention								
Mean DV	0.299	0.299	0.901	0.901	0.403	0.403	0.831	0.831
Std.Dev. DV	0.458	0.458	0.299	0.299	0.491	0.491	0.375	0.375

Notes: Table 3.2 shows the OLS estimates of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant i perceives the **fakes** on Corona vaccines and nutrition in Wave I and in Wave II of the survey on average as *Very credible*, *Credible* or *Indecisive*. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

comparing the MediaLiteracy to the NoIntervention group, respectively.[38]

Our first main result is that the fact-checking intervention reduces the credibility of fakes on Corona vaccines in Wave I of the survey. In Wave I, participants from the FactChecking group are on average 7 to 8 percentage points less likely to perceive fakes on Corona vaccines on average as credible than participants from the NoIntervention group. The estimate is statistically significant at the 1%-level, the effect size corresponds to about 15.5% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 23.7% of its mean value. In contrast to that, we find no statistically significant differences between the FactChecking and the NoIntervention group for fakes that are not corrected by fact-checks, i.e., fakes on Corona vaccines in Wave II of the survey and fakes on nutrition in either wave.

Second, we find that the media literacy intervention reduces the credibility of fakes more generally than fact-checking. In Wave I, participants from the MediaLiteracy group are about 10 percentage points less likely to consider fakes on Corona vaccines on average as credible than participants from the NoIntervention group. The estimate is statistically significant at the 1%-level; the effect size corresponds to about 22.9% of a standard deviation in the dependent variable and to 35% of its baseline value, whereby the effect is even larger than for the FactChecking group.[39] Unlike fact-checking, the media literacy intervention also reduces the credibility of all fakes on nutrition and of fakes on Corona vaccines in Wave II of the survey, although the latter effect is small and not statistically significant when we include our controls. The estimates for nutrition, however, correspond to about 18.1% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 8.2% of its mean value in both waves of the survey.[40]

Crucially, neither intervention reduces the credibility of facts (see Table B3 in Appendix B.4), i.e., participants do not become more skeptical towards social media postings per se. Instead, our results indicate that the fact-checking and, in particular, the media literacy intervention enhance participants' truth discernment (Pennycook and Rand, 2021), whereby they can better distinguish between false and correct

[38]Tables B7 to B9 in Appendix B.4 show the results for a direct comparison of the FactChecking and the MediaLiteracy group.

[39]The difference between the FactChecking and the MediaLiteracy group is not statistically significant, though (two-sided t-test, $p = 0.176$).

[40]Note that the mean perceived credibility of fakes on Corona vaccines is much lower than for fakes on nutrition. To analyse whether fact-checking more credible fakes enhances participants' critical evaluation of subsequent posts, we varied the exposure by presenting fakes with a higher mean baseline credibility in a follow-up study, which were fact-checked before participants proceeded to a new topic without further intervention. For a detailed discussion of the findings see Appendix B.5.3.

information that they encounter online.[41]

Truth discernment This result is further confirmed by an explicit analysis of truth discernment. In particular, we consider fakes and facts on Corona vaccines (nutrition) in Wave I (Wave II) of the survey in a single regression equation and estimate

$$cred_{ij} = \alpha_0 + \alpha_1 fake_j + \alpha_2 TG_i + \alpha_3 (TG_i * fake_j) + \theta X_i + e_{ij} \tag{3.2}$$

by OLS, where $cred_{ij}$ is a dummy equal to one if participant i perceived group j of news items (fakes or facts) on average as *Very credible*, *Credible* or *Indecisive*, TG_i is a treatment indicator with the NoIntervention group as omitted category, and $fake_j$ indicates fakes (in contrast to facts). The parameter α_1 thus corresponds to the average difference in the perceived credibility of fakes and facts – which we interpret as truth discernment – in the baseline NoIntervention group. The parameter of interest is α_3, which corresponds to the average difference in truth discernment between the NoIntervention and the FactChecking or the MediaLiteracy group, respectively.

Table B4 in Appendix B.4 displays our results. Panel A shows that relative to facts, participants from the FactChecking group are on average 6.8 percentage points less likely to perceive fakes on Corona vaccines as credible than participants from the NoIntervention group. Hence, fact-checking enhances participants' truth discernment for the fakes that are explicitly targeted by our intervention. In contrast to that, we do not observe a statistically significant difference in truth discernment between the FactChecking and the NoIntervention group for Corona vaccines in Wave II of the survey or for nutrition in either wave.

Unlike fact-checking, the media literacy intervention enhances participants' truth discernment throughout the entire survey (Panel B). In particular, relative to facts, participants from the MediaLiteracy group are on average between 6.6 and 9.8 percentage points less likely to perceive fakes on Corona vaccines on average as credible than participants from the NoIntervention group. With the exception

[41] Participants in the MediaLiteracy group might expect to see a certain proportion of fakes, leading to potential experimenter demand effects. However, participants did not know the total number of news items, preventing specific inferences about the total number of fakes they would encounter. Additionally, the results in Tables B4 and B6 in Appendix B.4 indicate that participants in the MediaLiteracy group improved in distinguishing false from correct news and in considering trustworthy elements in facts and untrustworthy elements in fakes. Further, Table B30 in Appendix B.5.3.6 shows that participants in the MediaLiteracy group in the follow-up experiment also spend more time reading the posts. This supports the idea that their judgments are based on enhanced skills rather than an expectation of a certain proportion of fakes. Finally, our list experiments in Section 3.6.6.3 show no evidence of greater experimenter demand effects in the MediaLiteracy group compared to other groups.

of Corona vaccines in Wave II of the survey, all estimates are highly statistically significant at the 1%-level.

3.4.1.2 Factual knowledge

Table 3.3 shows the regression results for participants' factual knowledge on the topics the fakes are dealing with. Again, Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively.

Consistent with the results on credibility, Panel A shows that the fact-checking intervention enhances participants' factual knowledge on Corona vaccines in Wave I of the survey. Specifically, responses by the FactChecking group are on average about 0.32 standard deviations closer to the correct answer than responses by the NoIntervention group; the effect is statistically significant at the 1%-level.[42] Somewhat surprisingly, we also observe that participants from the FactChecking group give better answers to the factual knowledge questions on nutrition in Wave I of the survey. According to our estimates, responses by the FactChecking group are about 0.08 standard deviations closer to the correct answer than responses by the NoIntervention group; the effect is statistically significant at the 5%-level. A potential explanation is that the presence of fact-checking makes participants generally more cautious towards implausible information, although it does not affect the perceived credibility of fakes that are not explicitly corrected. This would be in line with recent findings by Barrera et al. (2020) and Nyhan et al. (2020), who show that fact-checks can improve the accuracy of respondents' factual beliefs, but fail to affect more deep-rooted perceptions and attitudes (see Section 3.4.1.3 for further discussion). Consistent with that, we find no statistically significant differences in factual knowledge on fakes between the FactChecking and the NoIntervention group in Wave II of the survey, where no further fact-checks are shown.

Analogous to the results on credibility, Panel B shows that the media literacy intervention enhances participants' factual knowledge more generally than fact-checking. In Wave I of the survey, responses by the MediaLiteracy group are about 0.22 standard deviations closer to the correct answer than responses by the NoIntervention group for fakes on Corona vaccines, and about 0.08 standard deviations closer for fakes on nutrition. Both effects are statistically significant. The impact of the media literacy is thus smaller than the impact of the fact-checking intervention when the FactChecking group receives correct information

[42]Note that aggregating the standardized responses to Corona vaccine fakes, Corona vaccine facts, nutrition fakes, and nutrition facts in each wave of the survey causes the reported means and standard deviations in Table 3.3 to be unequal to zero and one, respectively.

in addition to the fakes, and roughly equivalent when it does not.[43] However, unlike fact-checking, the media literacy intervention could also improve participants' factual knowledge on Corona vaccines in Wave II of the survey. Specifically, responses by the MediaLiteracy group are about 0.14 standard deviations closer to the correct answer than responses by the NoIntervention group; the effect is statistically significant at the 1%-level when we include our controls. In contrast to that, we find no statistically significant difference in factual knowledge on nutrition between the MediaLiteracy and the NoIntervention group in Wave II of the survey, although the estimates have the expected sign. One plausible explanation is that, according to Table 3.1, fakes on nutrition seem to be more credible on average than fakes on Corona vaccines. As a result, the tips to spot false news could be more difficult to apply, which in turn entails a smaller difference between the MediaLiteracy and the NoIntervention group. In addition, the impact of our intervention is likely to decay over time (e.g., Maertens et al., 2021; Nyhan, 2021), which further reduces the effect size in Wave II of the survey.

Similar to the results on credibility, Table B5 in Appendix B.4 shows that neither intervention reduces participants' factual knowledge on topics that the facts are dealing with. Hence, both the fact-checking and the media literacy intervention enhance participants' factual knowledge on average. Moreover, Table B6 shows that the fact-checking intervention enhances participants' truth discernment for Corona vaccines in Wave I of the survey, while there is no statistically significant difference in truth discernment between the FactChecking and the NoIntervention group otherwise. The media literacy intervention, in contrast, enhances truth discernment more generally. In particular, all estimates have the expected sign and almost all of them are statistically significant.

3.4.1.3 Attitudes

Table 3.4 displays the regression results for participants' attitudes towards Corona vaccination and the intake of (needless) dietary supplements. Panel A displays the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively.

Panel A reveals that the impact of fact-checking on participants' attitudes is extremely limited. Although participants from the FactChecking group are more likely to state that they are *Very likely*, *Likely* or *Indecisive* to get vaccinated or

[43]The difference between the FactChecking and the MediaLiteracy group is statistically significant at the 5%-level for fakes on Corona vaccines in Wave I of the survey (two-sided t-test, $p = 0.034$) and weakly statistically significant at the 10%-level for fakes on nutrition in Wave II of the survey (two-sided t-test, $p = 0.097$).

Table 3.3: Distance to truth on topics covered by fakes

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.322	-0.319	-0.080	-0.080	-0.048	-0.051	0.021	0.030
	[0.047]	[0.048]	[0.035]	[0.035]	[0.054]	[0.054]	[0.038]	[0.038]
p-value	(0.000)	(0.000)	(0.023)	(0.024)	(0.371)	(0.350)	(0.596)	(0.426)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	-0.219	-0.220	-0.078	-0.081	-0.114	-0.138	-0.026	-0.041
	[0.048]	[0.048]	[0.037]	[0.036]	[0.053]	[0.052]	[0.038]	[0.037]
p-value	(0.000)	(0.000)	(0.032)	(0.027)	(0.031)	(0.009)	(0.488)	(0.264)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020
Baseline: No Intervention								
Mean DV	0.299	0.299	0.901	0.901	0.403	0.403	0.831	0.831
Std.Dev. DV	0.458	0.458	0.299	0.299	0.491	0.491	0.375	0.375

Notes: Table 3.3 compares distance to truth on topics that the Corona vaccine and nutrition **fakes** are dealing with between participants from the FactChecking (Panel A) and the MediaLiteracy (Panel B) and the NoIntervention group, respectively. All estimates are OLS estimates. The dependent variable is equal to participant i's average standardized distance to the correct answer. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

boostered against Covid-19 than participants from the NoIntervention group, much of the effect is driven by the smaller proportion of fully vaccinated participants in the latter (see Section 3.3.3). In particular, we find that a participant’s decision to get vaccinated in the past strongly predicts his or her intention to get vaccinated in the future. As a result, the estimated difference in the average willingness to get vaccinated between the FactChecking and the NoIntervention group shrinks and becomes statistically insignificant when we control for participants’ Corona vaccination status in columns 3 and 8.[44] Similarly, we do not find any statistically significant difference in participants’ willingness to consume dietary supplements in either wave of the survey. We must therefore conclude that the fact-checking intervention – though effective in reducing the perceived credibility of and enhancing factual knowledge on the fakes that are being targeted – fails to affect participants’ attitudes on average. This is consistent with earlier findings by Barrera et al. (2020), Swire et al. (2017), Nyhan et al. (2020), and Jerit and Zhao (2020), among others, who show that fact-checking can help to create “a more informed citizenry” (Nyhan et al., 2020, p.942), but struggles to change more deep-rooted perceptions and attitudes such as which political party to support or, as in our context, whether to get vaccinated against Covid-19 or not.

In line with the results on credibility and factual knowledge, Panel B shows that the media literacy intervention is effective in swaying participants’ attitudes. In particular, participants from the MediaLiteracy group are 3.4 to 4.8 percentage points more likely to state that they are *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19 than participants from the NoIntervention group. In contrast to fact-checking, this difference remains statistically significant when we control for participants’ Corona vaccination status.[45] The effect size corresponds to about 8.7% of a standard deviation in the dependent variable in the baseline NoIntervention group and 4.2% of its mean value in Wave I of the survey, and to about 11.6% of a standard deviation in the dependent variable and 6.1% of its baseline value in Wave II. Roughly 85% of all participants report that they are willing to get vaccinated or boostered against Covid-19, though. The relatively small effect size could thus be explained by a small proportion of participants who can still be convinced to get the shot; Section 3.4.3 computes persuasion rates à la DellaVigna and Kaplan (2007) to further address this issue. The estimates for participants’ willingness to consume needless dietary supplements are statistically insignificant.

[44]Note that the willingness to get vaccinated or boostered against Covid-19 is the only instance where the smaller proportion of fully vaccinated participants in the NoIntervention group plays a role.

[45]The differences between the FactChecking and the MediaLiteracy group are not statistically significant, though.

One potential explanation is that, unlike Corona vaccination, the intake of dietary supplements is typically based on year-long habits (Bailey et al., 2013), and even if the media literacy intervention could affect participants' attitudes, further effects from attitudes to habit change are typically modest (Verplanken and Orbell, 2022).

In sum, the results from Section 3.4.1 support the idea that the effect of fact-checking tends to be limited to the fakes that are being corrected, while enhancing participants' media literacy helps them to distinguish between fakes and facts more generally. Hence, in an environment where not every message can be fact-checked, media literacy interventions are likely to be more effective on average.

Table 3.4: Attitudes towards Corona vaccination and the intake of dietary supplements

Panel A: Fact-checking										
	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fact-checking	0.047	0.038	0.017	-0.001	0.003	0.057	0.048	0.034	0.011	0.021
	[0.023]	[0.021]	[0.019]	[0.028]	[0.028]	[0.026]	[0.025]	[0.022]	[0.031]	[0.031]
p-value	(0.037)	(0.068)	(0.341)	(0.959)	(0.908)	(0.032)	(0.049)	(0.127)	(0.717)	(0.504)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,225	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022	1,022
Panel B: Media literacy										
	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Media literacy	0.054	0.053	0.034	-0.041	-0.037	0.064	0.068	0.048	-0.022	-0.012
	[0.022]	[0.021]	[0.019]	[0.028]	[0.028]	[0.026]	[0.025]	[0.022]	[0.031]	[0.031]
p-value	(0.016)	(0.012)	(0.077)	(0.148)	(0.188)	(0.015)	(0.006)	(0.033)	(0.486)	(0.696)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,231	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020	1,020
Baseline: No Intervention										
Mean DV	0.813	0.813	0.813	0.548	0.548	0.781	0.781	0.781	0.552	0.552
Std.Dev. DV	0.390	0.390	0.390	0.500	0.500	0.413	0.413	0.413	0.497	0.497

Notes: Table 3.4 presents the OLS estimates of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19, or *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the near future. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 8 ("yes +"), we also control for participants' Corona vaccination status.

3.4.1.4 Multiple hypothesis testing

Our ITT analysis examines two treatments and twelve different outcome variables; i.e., we consider twenty-four hypotheses in sum. To take potential over-rejection of null hypotheses into account, we conduct a Romano-Wolf multiple hypothesis correction (Romano and Wolf, 2005). This procedure uses resampling methods (e.g., bootstrap) to control for the so-called familywise error rate, i.e., the probability of rejecting at least one true null hypothesis in the family of hypotheses under test. The Romano-Wolf procedure offers more power than the procedures by Bonferroni and Holm, and it is furthermore able to eliminate the so-called subset pivotality assumption from previous resampling-procedures such as Westfall-Young (Clarke et al., 2020).

Table 3.5 displays both the original and the Romano-Wolf p-values from our ITT analysis of the fact-checking (Panel A) and the media literacy intervention (Panel B), respectively. As expected, all p-values grow under the correction. Yet, our main results remain qualitatively intact in that the fact-checking intervention is limited to the fake news that are targeted, while the media literacy intervention helps more generally. In particular, we observe that the media literacy intervention has a statistically significant impact on the credibility of both fakes on Covid-19 vaccines and nutrition in Wave I of the survey even after the Romano-Wolf correction, and that one of the effects in Wave II survives.

Table 3.5: Romano-Wolf Multiple Hypothesis Correction

Panel A: Fact-checking												
	Wave I						**Wave II**					
	Corona			**Nutrition**			**Corona**			**Nutrition**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Original p-value	0.004	0.000	0.341	0.553	0.024	0.909	0.582	0.350	0.127	0.460	0.426	0.504
Romano-Wolf p-value	0.070	0.000	0.973	0.973	0.304	0.973	0.973	0.973	0.802	0.973	0.973	0.973
Panel B: Media literacy												
	Wave I						**Wave II**					
	Corona			**Nutrition**			**Corona**			**Nutrition**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Original p-value	0.000	0.000	0.077	0.001	0.027	0.188	0.143	0.009	0.033	0.006	0.264	0.700
Romano-Wolf p-value	0.001	0.000	0.642	0.028	0.323	0.878	0.816	0.134	0.354	0.097	0.948	0.973

Notes: Table 3.5 displays the original and the Romano-Wolf p-values from our ITT analysis. The Romano-Wolf p-values were obtained with the STATA package *rwolf2* (Clarke et al., 2020). Each specification includes the full set of control variables. In columns 3 and 9, we also include participants' vaccination status as additional control.

3.4.2 Heterogeneity in baseline beliefs

The average impact of our fact-checking and media literacy interventions is likely to depend on participants' prior beliefs. In particular, if participants update their beliefs in a Bayesian fashion, those with prior beliefs that are most different from the presented information would change their beliefs the most (see, e.g., Fong et al., 2024, for empirical evidence). This section demonstrates that our interventions are indeed more effective for supporters of the AfD ("Alternative for Germany"), a far-right populist party known for spreading misinformation on Corona vaccines (e.g., Gensing, 2021).[46] Specifically, we show that there is substantial effect heterogeneity between AfD and non-AfD supporters for fakes on Corona vaccines, but not for fakes on nutrition, where participants' prior beliefs are much more alike. Given that the effect heterogeneity is limited to fakes on Corona vaccines, we focus on that topic here, and defer the results on nutrition to Tables B10 and B11 in Appendix B.4.[47]

3.4.2.1 Fact-checking

Table 3.6 displays the average impact of fact-checking for AfD (Panel A) and non-AfD supporters (Panel B) on each of our main outcomes in each wave of the survey, respectively.

There are two main insights. First, each point estimate in Panel A is larger than its counterpart in Panel B, which means that the average impact of fact-checking is stronger for AfD than for non-AfD supporters. With the exception of factual knowledge, each of these differences is highly statistically significant (two-sided t-tests, $p < 0.001$). Second, there is ample heterogeneity in the baselines, i.e., the mean dependent variables in the NoIntervention group, which we interpret as participants' average prior beliefs: AfD supporters are on average almost twice as likely to perceive fakes on Corona vaccines as *Very credible*, *Credible* or *Indecisive* than non-AfD supporters, their responses to the factual knowledge questions are further away from the correct answer by roughly a third, and they are only half as likely to state that they are willing to get vaccinated or boostered against Covid-19. Hence,

[46] E.g., The AfD politician Björn Höcke claimed that Corona vaccines could cause infertility (`https://www.youtube.com/watch?v=X1Gw0k9CxUY&t=1346s`, viewed April 2024), Joachim Kuhs, also AfD politician, asserted that more people died from Corona vaccination in 2021 than people died from any kind of vaccination in the entire past twenty years (`https://www.youtube.com/watch?v=JmbLGTH4PDw`, viewed April 2024), and the AfD Bavaria denies that vaccines protect against severe progression of Covid-19 (`https://www.tagesschau.de/faktenfinder/afd-angst-impfungen-101.html`.)

[47] We do not find any systematic effect heterogeneity in terms of education, age, social media usage, support of policy measures to counteract the spread of the Corona virus, or prior knowledge on current events, health, and nutrition. When we compare participants who are fully vaccinated to those who are not, we find a similar, though less pronounced, pattern as for AfD supporters (see Section 3.6.6 for a detailed analysis). We preregistered a subgroup analysis that would exploit differences in participants' prior beliefs. While we initially thought that vaccination status or consent with the existing Corona remedies would be a good proxy, it turned out later that AfD support is seemingly even closer related to differences in prior beliefs. Hence, we preregistered the former but not the latter subgroup analysis. We also preregistered a subgroup analysis w.r.t. to time spent on social media, but, as argued, found no heterogeneous effects. See Appendix B for further details.

one explanation for the effect heterogeneity between AfD and non-AfD supporters is that the proportion of participants who can still update their beliefs is substantially larger for the former than for the latter group. We elaborate on this idea in Section 3.4.3, where we compute persuasion rates for each of our interventions. Differing trust in the fact-checks, in contrast, does not seem to drive the effect heterogeneity: AfD supporters perceive the fact-checks on Corona vaccines on average as slightly less credible as non-AfD supporters (see Section 3.6.4).

Column 3 shows that AfD supporters from the FactChecking group are 13.7 percentage points more likely to state that they are willing to get vaccinated or boostered than AfD supporters from the NoIntervention group. The effect is statistically significant at the 5%-level; it corresponds to 27.8% of a standard deviation in the dependent variable in the baseline NoIntervention group and to 35.1% of its mean value. This result is especially remarkable given that fact-checking typically fails to affect participants' attitudes (see Section 3.4.1.3). Here, the relatively strong impact of fact-checking on the credibility of and factual knowledge on fakes, combined with the initially small average proportion of AfD supporters who want to get vaccinated or boostered (39.0% vs. 82.5% for the non-AfD supporters), is potent enough to sway attitudes of AfD supporters, while attitudes of non-AfD supporters remain unaffected.

3.4.2.2 Media literacy

Analogous to Table 3.6, Table 3.7 displays the average impact of the media literacy intervention for AfD supporters (Panel A) and non-AfD supporters (Panel B) on each of our main outcomes in each wave of the survey, respectively. Again, the impact of our intervention is larger for AfD supporters than for non-AfD supporters; with the exception of factual knowledge, each of these differences is highly statistically significant (two-sided t-tests, $p < 0.001$). We also observe large differences in participants' baselines. Hence, part of the effect heterogeneity can be explained by the different proportion of participants who can still update their beliefs (see Section 3.4.3 for further discussion).

Similar to fact-checking, we find that the media literacy intervention has a large positive impact on AfD supporters' attitudes towards Corona vaccination. Specifically, AfD supporters from the MediaLiteracy group are about 14.9 percentage points more likely to state that they are *Very likely* or *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19 than AfD supporters from the NoIntervention group. The effect is statistically significant at the 5%-level; it corresponds to about 30.3% of a standard deviation in the dependent variable in the baseline NoIntervention group and to 38.2% of its mean value. In contrast to fact-checking, this difference remains

Table 3.6: Heterogeneity in baseline beliefs on Corona vaccination – FactChecking

Panel A: Fact-checking – AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.166	-0.362	0.137	-0.221	-0.325	0.090
	[0.099]	[0.171]	[0.062]	[0.096]	[0.178]	[0.081]
p-value	(0.095)	(0.036)	(0.029)	(0.024)	(0.072)	(0.267)
Controls	yes	yes	yes +	yes	yes	yes +
N	114	115	115	99	99	99
Baseline: No Intervention						
Mean DV	0.508	0.509	0.390	0.788	0.358	0.404
Std.Dev. DV	0.504	0.847	0.492	0.412	0.795	0.495
Panel B: Fact-checking – non-AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.055	-0.316	0.006	0.046	-0.032	0.030
	[0.025]	[0.051]	[0.019]	[0.031]	[0.058]	[0.023]
p-value	(0.029)	(0.000)	(0.752)	(0.145)	(0.576)	(0.192)
Controls	yes	yes	yes +	yes	yes	yes +
N	1,107	1,110	1,110	923	923	923
Baseline: No Intervention						
Mean DV	0.277	0.386	0.825	0.369	0.331	0.779
Std.Dev. DV	0.448	0.840	0.381	0.483	0.850	0.415

Notes: Table 3.6 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) for our **Fact-checking** intervention. In columns 1 and 4, the dependent variable is a dummy equal to one if participant i perceives the fakes on Corona vaccines as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant i's average standardized distance to the correct answer. In columns 3 and 6, the dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 6 ("yes +"), we also control for participants' Corona vaccination status.

statistically significant in Wave II of the survey, and the effect size is comparable to Wave I. This suggests that the media literacy intervention is even more successful in swaying participants' attitudes. Though, none of the differences in means between the FactChecking and the MediaLiteracy group are statistically significant when evaluated using a two-sided t-test.

In sum, we find that both the fact-checking and the media literacy intervention are more effective for AfD supporters, whose prior beliefs on Corona vaccines are on average further away from the truth and can thus be updated more strongly. This result stands in contrast to previous findings on motivated reasoning (e.g., Jerit and Zhao, 2020; Lewandowsky et al., 2012), whereby preexisting worldviews or attachments to a political party can impede efforts to debunk fake news. However, Pennycook et al. (2021) and Pennycook et al. (2020) argue that it is often a lack of attention rather than partisanship that drives such results. Our evidence is consistent with the latter line of thought (see also Section 3.5), since both the fact-checking and the media literacy intervention increase the awareness of fake news on Corona vaccines, and thereby induce participants to update their beliefs.

3.4.3 Persuasion rates

Sections 3.4.1 and 3.4.2 reveal that many participants are willing to get vaccinated or boostered against Covid-19 irrespective of the interventions. As a result, the ITT effects are relatively small. To adjust our estimates for the share of participants left to be convinced, we compute persuasion rates à la DellaVigna and Kaplan (2007), both for the full sample as well as for AfD and non-AfD supporters, respectively. To that end, we define the share of participants left to be convinced as the NoIntervention group's share of participants stating to be *Very unlikely* or *Unlikely* to get vaccinated or boostered on the 5-point Likert scale.[48] Then, we divide the ITT estimates from our preferred specifications ("yes +") in Tables 3.4, 3.6, and 3.7 by that share.

Table 3.8 shows that around 21.7% (25.9%) of our participants could still be persuaded in Wave I (Wave II) of the survey. Specifically, 61% of the AfD supporters and 17.5% of the non-AfD supporters could still be persuaded in Wave I, and 59.6% (22.1%) in Wave II. Hence, the fact-checking intervention could convince about 7.8% (13.1%) of all persuadable participants in Wave I (Wave II) of the survey, whereas the media literacy intervention could convince 15.7% in Wave I and 18.5% in Wave II. These magnitudes are similar to those in comparable papers (e.g., Barrera et al., 2020, p.13). The persuasion rates for AfD supporters are considerably larger than the

[48] We obtain similar results using a cutoff here we define the share of participants left to be convinced as the NoIntervention group's share of participants stating to be *Very unlikely*, *Unlikely*, or *Indecisive* to get vaccinated or boostered.

Table 3.7: Heterogeneity in baseline beliefs on Corona vaccination – MEDIALITERACY

Panel A: Media literacy – AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	-0.134	-0.279	0.149	-0.161	0.090	0.136
	[0.096]	[0.166]	[0.071]	[0.091]	[0.170]	[0.068]
p-value	(0.167)	(0.095)	(0.039)	(0.080)	(0.600)	(0.047)
Controls	yes	yes	yes +	yes	yes	yes +
N	116	116	116	101	101	101
Baseline						
Mean DV	0.508	0.509	0.390	0.788	0.358	0.404
Std.Dev. DV	0.504	0.847	0.492	0.412	0.795	0.495
Panel B: Media literacy – non-AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	-0.096	-0.208	0.021	-0.027	-0.151	0.042
	[0.024]	[0.050]	[0.020]	[0.031]	[0.056]	[0.023]
p-value	(0.000)	(0.000)	(0.280)	(0.379)	(0.007)	(0.075)
Controls	yes	yes	yes +	yes	yes	yes +
N	1,115	1,115	1,115	919	919	919
Baseline						
Mean DV	0.277	0.386	0.825	0.369	0.331	0.779
Std.Dev. DV	0.448	0.840	0.381	0.483	0.850	0.415

Notes: Table 3.7 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) for our **Media literacy** intervention. In columns 1 and 4, the dependent variable is a dummy equal to one if participant i perceives the fakes on Corona vaccines as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant i's average standardized distance to the correct answer. In columns 3 and 6, the dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 6 ("yes +"), we also control for participants' Corona vaccination status.

persuasion rates for non-AfD supporters. Hence, even if we account for differences in the share of participants left to be convinced, both interventions are more effective for AfD than for non-AfD supporters.[49]

A complementary explanation for our finding could be that it is *ceteris paribus* easier to convince the former. E.g., DellaVigna and Gentzkow (2010) argue that persuasion is more effective when receivers of novel information are less certain about the truth (p.654). Similarly, Kuklinski et al. (2000) distinguish between *mis*informed – those who have wrong beliefs and hold them firmly (p.792) – and *un*informed citizens. Consistent with that, we find that AfD supporters are on average less certain about their prior knowledge on current events, health, and nutrition. In particular, when we regress the uncertainty indicator from Section 3.3.4 on the AfD dummy, the resulting estimate indicates that AfD supporters are on average 4.1 percentage points more likely to be uncertain about their prior knowledge than non-AfD supporters; the effect is statistically significant at the 5%-level.[50] Similarly, when we further divide the sample of AfD supporters into those who are uncertain about their prior knowledge and those who are not, we find that the ITT estimates are larger and more statistically significant for the former group. Hence, the effect heterogeneity between AfD and non-AfD supporters is likely to be driven both by wrong priors and by uncertainty about them. Our results thus show that it is important not to give up on those with seemingly extreme opinions; rather, one should try to address exactly those people.

3.5 Mechanisms

Section 3.4.1 shows that the effectiveness of fact-checking tends to be limited to the fakes that are being corrected, while the media literacy intervention helps to distinguish between fakes and facts more generally, both immediately and in the short-run. A reasonable explanation is that the media literacy intervention raises participants' attention towards the Facebook postings and helps them to critically evaluate the postings' accuracy. Fact-checking, in contrast, turns participants into passive recipients of the specific corrections and thus fails to markedly enhance their skills.

To support the plausibility of this mechanism, this section shows that participants from the MediaLiteracy group are on average more likely to actively search

[49]See Tables B14 - B12 in Appendix B.4 for persuasion rates on our other binary outcomes.

[50]In contrast to that, there is no evidence that AfD supporters' prior knowledge on current events, health, and nutrition is worse than that of non-AfD supporters. The proportions of AfD and non-AfD supporters who are on the edge of being persuaded (i.e., stating to be *Undecided*) to get vaccinated or boostered are similar, too.

Table 3.8: Persuasion rates Corona vaccination

Panel A: Fact-checking						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Persuasion rate	0.078	0.225	0.034	0.131	0.151	0.136
Share to be persuaded	0.217	0.610	0.175	0.259	0.596	0.221
Panel B: Media literacy						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Persuasion rate	0.157	0.244	0.120	0.185	0.228	0.190
Share to be persuaded	0.217	0.610	0.175	0.259	0.596	0.221

Notes: Table 3.8 displays the persuasion rates for our fact-checking (Panel A) and media literacy interventions (Panel B) for Wave I and Wave II of the survey, respectively. Columns 1 and 4 consider all participants in the NoIntervention, FactChecking, and MediaLiteracy groups. Columns 2 and 5 consider only AfD supporters, columns 3 and 6 only non-AfD supporters. The *Share to be persuaded* corresponds to the proportion of participants in the NoIntervention group stating to be *Very unlikely* or *Unlikely* to get vaccinated or boostered.

for further information than participants from the NoIntervention group. In addition, they become better at considering untrustworthy elements in fakes and trustworthy elements in facts. Figures B11 to B13 in Appendix B.3 illustrate our results; further details are given below. In Appendix B.5.3.6, we present results from a follow-up survey demonstrating that participants from the MediaLiteracy group spent significantly more time with the fakes and facts than participants from the NoIntervention group. Participants from the FactChecking group, in contrast, spent significantly less time with the postings. We interpret this as further evidence that the media literacy intervention raises awareness and critical evaluation, while fact-checking does not achieve this effect.

3.5.1 Search for further information

We first examine if participants actively search for further information. To this end, we generate a dummy equal to one if participant i reports to have used the Internet to answer the factual knowledge questions on Corona vaccines and nutrition, and use this dummy as dependent variable in equation (3.1).[51]

Consistent with the proposed mechanism, Table 3.9 shows that all estimates for the FactChecking group are negative, but they are not statistically significant

[51] We did not pre-register the question on search for further information.

(Panel A). In contrast to that, all estimates for the MediaLiteracy group are positive, and they are statistically significant at the 10%-level in Wave I of the survey (Panel B). Specifically, participants from the MediaLiteracy group are 4.7 to 4.9 percentage points more likely to search for further information than participants from NoIntervention group. The effect size corresponds to 9.8% of a standard deviation in the dependent variable in the baseline NoIntervention group.

Table 3.9: Search for further information

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.041	-0.034	-0.037	-0.035	-0.015	-0.008	-0.032	-0.026
	[0.028]	[0.027]	[0.025]	[0.025]	[0.030]	[0.030]	[0.028]	[0.028]
p-value	(0.144)	(0.208)	(0.150)	(0.165)	(0.628)	(0.777)	(0.261)	(0.338)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	0.049	0.048	0.047	0.047	0.010	0.016	0.036	0.043
	[0.028]	[0.028]	[0.027]	[0.026]	[0.030]	[0.030]	[0.029]	[0.029]
p-value	(0.083)	(0.085)	(0.083)	(0.079)	(0.733)	(0.586)	(0.215)	(0.136)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020
Baseline: No Intervention								
Mean DV	0.400	0.400	0.307	0.307	0.361	0.361	0.295	0.295
Std.Dev. DV	0.490	0.490	0.462	0.462	0.481	0.481	0.456	0.456

Notes: Table 3.9 shows the OLS estimates of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant i reports to have used the Internet to respond to the factual knowledge questions (fakes and facts) on Corona vaccines and nutrition in Wave I and in Wave II, respectively. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

3.5.2 Considering (un-)trustworthy elements in fakes and facts

Next, we show that the media literacy intervention helps participants to consider untrustworthy elements in fakes and trustworthy elements in facts, whereby they

Table 3.10: Absolute number of untrustworthy elements considered in fakes

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.023	-0.093	-0.048	-0.085	0.070	-0.022	0.014	-0.041
	[0.190]	[0.184]	[0.099]	[0.097]	[0.264]	[0.256]	[0.138]	[0.135]
p-value	(0.902)	(0.610)	(0.627)	(0.383)	(0.792)	(0.931)	(0.917)	(0.759)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,030	1,030	1,030	1,030
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	1.242	1.194	0.679	0.660	1.088	0.903	0.725	0.680
	[0.210]	[0.197]	[0.109]	[0.106]	[0.290]	[0.271]	[0.163]	[0.155]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026
Baseline: No Intervention								
Mean DV	3.249	3.249	1.256	1.256	4.469	4.469	1.651	1.651
Std.Dev. DV	3.293	3.293	1.714	1.714	4.144	4.144	2.337	2.337

Notes: Table 3.10 compares the absolute number of elements considered as untrustworthy in fakes on Corona vaccines and nutrition for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B) in Wave I and Wave II of the survey, respectively. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

can better distinguish between false and correct information that they encounter online. To that end, we separately compute how many elements (in absolute terms) participants consider as trustworthy and untrustworthy, both in fakes and in facts.[52] For all of these measures we compare participants from the NoIntervention to the FactChecking and to the MediaLiteracy group.[53]

Table 3.10 confirms that the media literacy intervention induces participants to consider a larger absolute number of elements in fakes as untrustworthy, while the fact-checking intervention has no such effect. In particular, all estimates for fact-checking are close to zero and statistically insignificant (Panel A), while the estimates for the media literacy intervention are positive and statistically significant at the 1%-level (Panel B). In Wave I of the survey, participants from the MediaLiteracy group consider on average 1.2 more elements in fakes on Corona vaccines

[52] For more details on how the measure is constructed see Section 3.3.4, a detailed description of how and which elements can be clicked is available in Section 3.3.1.1.

[53] We did not pre-register the consideration of (un-)trustworthy elements.

Table 3.11: Absolute number of trustworthy elements considered in facts

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	0.075	0.026	0.011	0.009	0.042	0.023	0.077	0.051
	[0.167]	[0.162]	[0.187]	[0.184]	[0.138]	[0.136]	[0.180]	[0.181]
p-value	(0.653)	(0.872)	(0.953)	(0.962)	(0.763)	(0.867)	(0.670)	(0.777)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,030	1,030	1,030	1,030
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	1.059	0.998	0.741	0.672	0.873	0.809	0.758	0.686
	[0.196]	[0.185]	[0.200]	[0.193]	[0.158]	[0.149]	[0.189]	[0.182]
p-value	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026
Baseline: No Intervention								
Mean DV	2.607	2.607	3.239	3.239	2.231	2.231	3.012	3.012
Std.Dev. DV	2.975	2.975	3.205	3.205	2.103	2.103	2.711	2.711

Notes: Table 3.11 compares the absolute number of elements considered as trustworthy in facts on Corona vaccines and nutrition for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B) in Wave I and Wave II of the survey, respectively. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

and 0.7 more elements in fakes on nutrition as untrustworthy than participants from the NoIntervention group.[54] For fakes on Corona vaccines, the effect size corresponds to 36.3% of a standard deviation in the dependent variable in the baseline NoIntervention group and to 36.7% of its mean value. For fakes on nutrition, the effect size corresponds to 38.5% of a standard deviation in the dependent variable in the baseline NoIntervention group and to 52.5% of its mean value.[55] The intervention's impact persists in Wave II of the survey. Specifically, participants from the MediaLiteracy group consider on average 0.9 more elements in fakes on Corona vaccines and 0.7 more elements in fakes on nutrition as untrustworthy than

[54]Table B17 in Appendix B.4 reports exemplary results for one specific type of elements: *Verified account*. In line with the results from this section, we find that participants from the MediaLiteracy group are more likely to consider a *Verified account* label in facts on Covid-19 and nutrition as trustworthy than participants in the NoIntervention group. The fact-checking intervention, in contrast, has no such effect.

[55]All differences between the FactChecking and the MediaLiteracy group are statistically significant at the 1%-level (two-sided t-tests, $p < 0.000$).

participants from the NoIntervention group. For fakes on Corona vaccines, this corresponds to 21.7% of a standard deviation in the dependent variable and to 20.1% of its mean value. For fakes on nutrition, this corresponds to 29.1% of a standard deviation in the dependent variable and to 41.2% of its mean value.

Analogously, Table 3.11 confirms that the media literacy intervention induces participants to consider a larger number of elements in facts as trustworthy, while the fact-checking intervention has no such effect. In particular, all estimates for fact-checking are close to zero and statistically insignificant (Panel A), while the estimates for the media literacy intervention are positive and statistically significant at the 1%-level (Panel B). In Wave I of the survey, participants from the MediaLiteracy group consider on average 1 more element in facts on Corona vaccines and 0.7 more elements in facts on nutrition as trustworthy than participants from the NoIntervention group. For facts on Corona vaccines, the effect corresponds to 33.5% of a standard deviation in the respective dependent variable in the baseline NoIntervention group and to 38.3% of its mean value. For facts on nutrition, the effect corresponds to 20.9% of a standard deviation in the respective dependent variable in the baseline NoIntervention group and to 20.7% of its mean value. Again, the impact of the media literacy intervention persists in Wave II of the survey. Specifically, participants from the MediaLiteracy group consider on average 0.8 more elements in facts on Corona vaccines and 0.7 more elements in facts on nutrition as trustworthy than participants from the NoIntervention group. For facts on Corona vaccines, this corresponds to 38.5% of a standard deviation in the dependent variable and to 36.3% of its mean value. For facts on nutrition, this corresponds to 25.3% of a standard deviation in the dependent variable and to 22.8% of its mean value.

Crucially, participants from the MediaLiteracy group do not generally consider a larger number of elements as (un-)trustworthy in the fakes and facts. In particular, the media literacy intervention does not increase the number of elements considered as trustworthy in fakes (Table B15 in Appendix B.4), and its impact on the number of elements considered as untrustworthy in facts is small and limited to (arguably untrustworthy) emojis in facts on nutrition (Table B16 in Appendix B.4). Hence, consistent with the results from Section 3.4.1, the media literacy intervention does not make participants more skeptical towards social media postings per se, but rather helps them to distinguish between trustworthy and untrustworthy (elements of the) information.[56]

[56]Note that our primary objective is to assess whether participants in the MediaLiteracy group improve in distinguishing trustworthy and untrustworthy elements within a post, rather than evaluating the post’s overall accuracy, which likely depends on the interplay of all elements.

3.6 Further analyses

This section provides further context for the interpretation of our main results. Specifically, we examine the PassiveControl group to study if our interventions can offset the damage that fake news are causing, we consider the JustFacts group to further investigate the limited effectiveness of fact-checking, we discuss potential heterogeneity in the fakes and fact-checks that we show to our participants, and we provide a battery of robustness checks for our main results.[57]

3.6.1 Comparison to PassiveControl group

The main purpose of our paper is to study whether and to what extent fact-checking and media literacy interventions are able to debunk fake news that circulate on social media. The relevant benchmark are thus participants from the NoIntervention group, who are exposed to fakes and facts without further intervention. While this comparison illustrates the effectiveness of our interventions in an environment where all participants see fakes and facts, it does not uncover to what extent the interventions are able to reverse the harm the fakes are causing. To better interpret the magnitude of our main estimates in that regard, this section compares responses from the PassiveControl group – i.e., participants who did not see any fakes or facts from Facebook at all – to the FactChecking and the MediaLiteracy group, respectively. The smaller the difference between those groups, the more effective is the respective intervention in repealing the impact of fakes.[58] We complement the analysis with a comparison of the PassiveControl to the NoIntervention group, which provides a benchmark for the absolute impact of fakes. Figures B9 and B10 in Appendix B.3 illustrate our results; further details are provided below.[59]

3.6.1.1 Factual knowledge

Table B18 in Appendix B.4 reveals that exposure to fakes on Corona vaccines and nutrition substantially impairs participants' factual knowledge, and that neither the fact-checking nor the media literacy intervention can fully offset the effect. In particular, we find that (almost) all estimates are positive and statistically significant at the 1%-level, which means that responses by the FactChecking, the

[57] We preregistered our analyses with a PassiveControl and a JustFacts group; see Appendix B.2 for details.

[58] Our interventions do not overcompensate the impact of fakes.

[59] Recall that participants in the PassiveControl group are not asked to rate the credibility of fakes and facts (see Section 3.3.1), since the idea of the PassiveControl group is to measure factual knowledge and attitudes when not being exposed to any of the fakes or facts. Therefore, we cannot compare the credibility outcomes measured in the main treatments with the PassiveControl group.

MediaLiteracy, and the NoIntervention group are significantly further away from the correct answer than responses by the PassiveControl group.

There are two additional insights. First, in comparison to the absolute impact of fakes (Panel C), the average effectiveness of our interventions is relatively small. E.g., in Wave I of the survey, the fact-checking intervention repeals less than 40% of the damage caused by fakes on Corona vaccines, and 80% of the damage caused by fakes on nutrition. Similarly, the media literacy intervention repeals just 26.2% of the damage caused by fakes on Corona vaccines, and 80% for nutrition. Thus, while both interventions improve factual knowledge relative to the NoIntervention group – i.e., in an environment where all participants see fakes and facts – they do not reverse the harm of fakes entirely.

Second, the estimates for nutrition are smaller than the estimates for Corona vaccines, i.e., responses are on average more similar to the PassiveControl group. A plausible explanation could be that the fakes on Corona vaccines are more persuasive than the fakes on nutrition, whereby participants' factual knowledge on the former is shifted further away from their prior than their factual knowledge on the latter.

3.6.1.2 Attitudes

In contrast to factual knowledge, Table B19 in Appendix B.4 shows that exposure to fakes has a relatively small, if any, impact on participants' attitudes (Panel C), and that both the fact-checking and, in particular, the media literacy intervention can effectively repeal that impact. Specifically, we find that (almost) all estimates are close to zero and statistically insignificant, which means that participants from the FactChecking, the MediaLiteracy, and the NoIntervention group report a similar willingness to get vaccinated (or boostered) against Covid-19 as well as a similar willingness to consume (needless) dietary supplements as participants from the PassiveControl group. This confirms the results from Section 3.4, whereby the majority of participants wants to get vaccinated or boostered irrespective of the interventions, and whereby it is difficult to affect habit-based attitudes on dietary supplements. Hence, the modest impact of our interventions on participants' attitudes (see Table 3.4) can also be explained by the small absolute impact of fakes: if exposure to fakes does not sway participants' attitudes to begin with, there is nothing that the fact-checking or the media literacy intervention could change.[60]

[60]This is in contrast to Barrera et al. (2020), who find that exposure to fake news is highly persuasive. However, Barrera et al. (2020) consider fake news on migration in France, while we consider fake news on Corona vaccines and nutrition. The diverging results could thus be driven by differences in the context of the fakes, i.e., it could be easier to sway participants' voting intentions than their attitudes on Corona vaccination and dietary supplements.

3.6.2 Comparison to JustFacts group

Section 3.5 demonstrates that the media literacy intervention helps participants to better distinguish between fakes and facts, while fact-checking fails to markedly enhance their skills. A complementary explanation for the smaller effectiveness of fact-checking is that the corrections often repeat false claims and thus induce "anchoring" (Tversky and Kahneman, 1974) or "continued influence effects" (Lewandowsky et al., 2012), whereby users' beliefs are biased towards the initially presented values. To study the role of such effects in our context, this Section compares factual knowledge of the JustFacts group – i.e., participants who only saw facts and fact-checks – to the NoIntervention and the PassiveControl group, respectively. If anchoring effects do not play a role, participants from the JustFacts group should have better factual knowledge than participants from the NoIntervention and the PassiveControl group, because they are given the correct information. If anchoring effects exist, however, we expect the JustFacts group's factual knowledge to be in between the NoIntervention and the PassiveControl group. The closer their responses are to the former, the stronger are the anchoring effects. Figure B9 in Appendix B.3 illustrates our results, further details are discussed below.

Table B20 in Appendix B.4 reveals that participants from the JustFacts have better factual knowledge on nutrition than participants from the PassiveControl, and better factual knowledge on both Corona vaccines and nutrition than participants from the NoIntervention group.[61] All differences are statistically significant at the 1%- or at the 5%-level.

There are three potential explanations for these results that are not mutually exclusive. First, consistent with Tversky and Kahneman (1974) and Lewandowsky et al. (2012), repetition of the false claims could stick in participants' memory. In particular, while all fact-checks on Corona vaccines restate the respective fake news, the fact-checks on nutrition do not recast any false or misleading numbers. As a result, fact-checking increases participants' factual knowledge on nutrition (as measured by the PassiveControl group), but reduces factual knowledge on Corona vaccines (although the JustFacts group still performs significantly better than the NoIntervention group).[62] Second, Section 3.6.4 reveals that participants from the JustFacts group perceive the fact-checks on nutrition as significantly more credible than the fact-checks on Corona vaccines. Thus, it could be that the former exhibit a stronger impact on participants' priors, whereby they update their beliefs more

[61] Due to a technical issue with one of the fact-checks on nutrition in Wave I of the survey, we do not aggregate participants' responses to the factual knowledge questions but consider just the one functioning fact-check instead.

[62] This explanation would also be consistent with the salience effects documented by Barrera et al. (2020).

extensively. Third and relatedly, participants' prior beliefs on nutrition could be less firm than their beliefs on Corona vaccines, and thereby more easy to sway. Similarly, as discussed in Section 3.6.1 above, their priors on nutrition could be worse than their priors on Corona vaccines (i.e., further away from the truth), leaving more room for improvement through the fact-checks. In sum, our evidence is in line with "anchoring" or "continued influence effects" that constitute one potential drawback of fact-checking, but we cannot exclude alternative explanations, either.

3.6.3 Heterogeneity in credibility of fakes

The effectiveness of our interventions is likely to depend on the specific fakes that we select for the experiment. To explore this issue in more detail, Figure B14 in Appendix B.3 displays the (disaggregated) mean credibility of all fakes as given by the NoIntervention group on a 5-point Likert Scale.

We find that there is substantial heterogeneity in the credibility of fakes. In particular, participants perceive the fakes on Corona vaccines on average as less credible than the fakes on nutrition. Moreover, there is heterogeneity within topics: the perceived credibility of fakes on Corona vaccines ranges from 1.65 to 2.28 and of fakes on nutrition from 2.75 to 3.41.

To further examine how fake credibility affects intervention effectiveness, we conduct a follow-up experiment where the NoIntervention group rates the fakes encountered during the first part of the survey as more credible than in our original study. Interestingly, in contrast to our original study, we detect a small but statistically significant effect of fact-checking on non-fact-checked posts in this follow-up. Fact-checks targeting less obvious fakes may thus lead readers to approach subsequent posts with greater criticism. Moreover, the findings of the follow-up strongly support our interpretation that the media literacy intervention enhances skills that can be more broadly applied than fact-checking. A detailed description of the follow-up is provided in Appendix B.5. For a direct comparison of intervention effects on nutrition posts in the original and follow-up experiments, see Table B29 in Appendix B.5.

3.6.4 Heterogeneity in credibility of fact-checks

According to Jerit and Zhao (2020), trust in the authors of corrective messages is a crucial cause for their effectiveness. In our context, distrust in the (authors of the) fact-checks could further explain why the fact-checking is less effective than the media literacy intervention. To explore the plausibility of this explanation, Figure B15 in Appendix B.3 displays the (disaggregated) mean credibility of all fact-checks on a 5-point Likert Scale.[63]

We find that the mean credibility for fact-checks on Corona vaccines in Wave I of the survey – i.e., those that were displayed to the FactChecking group – is surprisingly low: the fact-checks rate between 2.62 and 2.73 for participants of the FactChecking, and between 2.81 and 2.88 for participants of the JustFacts group. This is significantly less than for fact-checks on Corona vaccines in Wave II of the survey (ratings between 3.15 and 3.27) and for fact-checks on nutrition in either Wave (ratings between 3.60 and 3.79). One possible driver of these differences could be heterogeneity in the source. E.g., while both fact-checks on Corona vaccines in Wave II of the survey are released by *dpa*, Germany’s most renowned news wire, the fact-checks on Corona vaccines in Wave I stem from *Correctiv* and *AFP*, respectively. Although these are generally considered as reliable fact-checking organizations, they might be less known among the participants of our experiment, and thus perceived as less trustworthy. On the other hand, one of the fact-checks on nutrition in Wave II of the survey comes “just” from an online platform and three from national public authorities, but Figure B15 shows that there are just minor differences in their perceived credibility. Hence, while it seems plausible that small trust in fact-checking contributes to its relative ineffectiveness, we cannot claim with certainty that the authors of the fact-checks are crucial components in this.

How does this affect the interpretation of our results? While the existing literature agrees that trust in fact-checks is a crucial cause for their effectiveness, it also shows that little trust in fact-checks does not necessarily mean that fact-checking does not work at all: E.g., Martel and Rand (2024) find that while warning effects are somewhat smaller among participants with lower trust in the fact-checkers, the warning labels still significantly reduced belief in false information (by 12.9%, p.1960), even among those most skeptical of fact-checkers. Similarly, X. Liu et al. (2023) observed few differences in effectiveness across various fact-checking sources, although sources deemed more credible showed greater effectiveness.

These findings suggest that the effects we observe regarding the efficacy of fact-checking on the perceived credibility of fakes on Covid-19 vaccines likely represent lower bounds. The effects could potentially be even stronger if trust in the fact-checks was higher.

3.6.5 Order of fact-checks

In our main experiment fact-checks were displayed to the participants before seeing the actual fake post. In reality however, platforms implement fact-checks both

[63] Recall that the FactChecking group was only shown fact-checks on Corona vaccines in Wave I, while the JustFacts group saw fact-checks on all topics in both waves of the survey.

before and after posts with fake content. We randomized this order in the follow-up experiment to test whether the order of fact-check and fake influences the initial experiment's findings. A comparison of the effectiveness of fact-checking for the different orders within the FactChecking group suggests weak evidence that fact-checking is more effective when fact-checks are presented after the fake. However, this result is not robust to a slightly stricter cutoff in the dependent variable (while the findings in our original experiment are robust to different cutoffs).

As we observe only small and not very robust differences for the order of the fact-check, we conclude that the influence of the order is most likely not the driving force behind our main results. If the order of the fact-check played a role in our main experiment, the estimates we obtained there should be regarded as lower bounds. Moreover, the qualitative nature of our findings with respect to the MediaLiteracy group remains unchanged: Even when fact-checks are displayed after the fake the effect of the fact-checking intervention is substantially lower than that of the media literacy intervention. A detailed description of this analysis and regression results are denoted in Appendix B.5.3.4.

3.6.6 Robustness checks

This section provides several analyses that support the robustness of our main results. In particular, we show that heterogeneity in terms of participants' vaccination status and the time span between Waves I and II does not play a role, we present evidence from list experiments that support the validity of the participants' self-reported attitudes, and we demonstrate that our results are robust to using an IV approach, where we take potential non-compliance into account.

3.6.6.1 Heterogeneity in terms of vaccination status

At the time of our experiment (September/October 2021), every German adult could be fully vaccinated against Covid-19 (two injections); yet, only about 81.8% of our participants reported to have taken the opportunity. This raises two potential concerns. First, participants who selected themselves into vaccination could be systematically different from those who did not, especially with respect to the questions on Corona vaccines. Second, we elicited participants' attitudes towards Corona vaccination with two different questions – "willingness to get vaccinated" vs. "willingness to get boostered" on a 5-point Likert-scale, respectively (see Section 3.3)– and responses to those two questions might not be entirely comparable. Our main analyses address these concerns with robustness checks, where we add participants' Corona vaccination status as an additional control (see Section 3.4). To further support our main findings, this section shows that they are robust to splitting the sample into participants who are fully vaccinated and those who are not.[64]

Consistent with the robustness checks that we already conducted, Tables B23 and B24 in Appendix B.4 show that the effect of our fact-checking and media literacy interventions are similar for participants who are fully vaccinated and those who are not. The ITTs of fact-checking (Table B23), for instance, hardly differ between the two subsamples; the main difference is the reduced precision in Panel B, stemming from the small sample of non-vaccinated participants. Table B24 reveals that the effect on the willingness to get vaccinated is larger for non-vaccinated than fully vaccinated participants in the MediaLiteracy group. This result is comparable to our findings on effect heterogeneity for AfD and non-AfD supporters from Section 3.4.2. It likely arises from differences in the baseline (i.e., there are more non-vaccinated participants left to be convinced) and, to some degree, from the differently posed question. In sum, however, we conclude that heterogeneity between participants who are fully vaccinated and those who are not is at most a minor concern.

3.6.6.2 Delay between Waves I and II

As we describe in Section 3.3.2, all participants received a re-invitation to Wave II about one week after they completed Wave I of the survey. However, as they could re-start the survey at any time after that, the delay between actual participation in the two waves is quite heterogeneous and lies between 8 and 45 days.[65]

We conduct two analyses to confirm that heterogeneity in the delay between Waves I and II is unlikely to affect our results. First, when we regress the number of days between participation in the two waves on our treatment indicators (with the NoIntervention group as omitted category) plus the full set of controls, almost all estimates are close to zero and statistically insignificant.[66] Hence, our treatments do not seem to influence when participants re-start the survey. Second, there is no evidence that differences in the delay between Waves I and II affect the magnitude of our main estimates. When we interact the number of days between participation in the two waves with the treatment indicators in equation (3.1), almost all interaction terms are close to zero and statistically insignificant. Similarly, when we split the

[64] For brevity, we only report the results for questions related to Corona vaccines. We do not find any systematic effect heterogeneity between participants who are fully vaccinated and those who are not with respect to questions on nutrition.

[65] We did not preregister the analysis of the delay between Waves I and II.

[66] The estimate for the PassiveControl group is negative and weakly statistically significant at the 10%-level, indicating that participants re-started the survey about half a day earlier than participants from the NoIntervention group.

sample at the median delay (= 15 days) and estimate equation (3.1) on the two subsamples, respectively, the resulting point estimates resemble those from Section 3.4.1. Note, however, that the delay between actual participation in Wave I and II is endogenous, so all robustness checks from this section should be interpreted with caution.

To interpret these results, note that a two-week delay is relatively short compared to other education intervention studies, where effects often persist for months (see Kaiser et al., 2022, for an overview). Thus, it is unsurprising that the effects are similar for participants who completed the survey at just slightly different times. Our findings also align with the literature on correcting misinformation. For instance, A. M. Guess et al. (2020) found that the effects of their media literacy intervention remained measurable and statistically significant after several weeks. Similarly, Carey et al. (2022) studied the decay in fact-checking efficacy and, consistent with our results, found no effect approximately two weeks post-intervention.

3.6.6.3 List experiments and experimenter demand bias

Participants' self-reported attitudes on Corona vaccines and dietary supplements might suffer from social desirability or experimenter demand bias if the participants anticipate that we as researchers are in favor of vaccination and against the consumption of needless dietary supplements. However, recent empirical evidence from both the behavioral economics as well as the political science literature suggests that experimenter demand biases are small, if they exist at all (De Quidt et al., 2018; Mummolo and Peterson, 2019). As argued in Section 3.3, we further minimize the risk of experimenter demand bias by *not* incentivizing the corresponding questions with a potential bonus payment. In addition to that, this section conducts two list experiments à la Blair and Imai (2012) – one for Corona vaccines, one for nutrition – to confirm that our participants do not conceal any socially undesirable opinions and attitudes.

For each list experiment, we randomly partition our participants into two groups. One group receives a list of five, the other group a list of six statements in random order, where the additional sixth statement is "I prefer not to get vaccinated against Covid-19." ("I take dietary supplements.") and the other five statements are about unrelated topics (see Appendix B.1.2 for the full lists). Then, we ask each participant how many of those statements he or she would agree with. Finally, we compute the difference in means between the two groups for the number of supported statements, which can be interpreted as the proportion of participants who indirectly concede that they do not want to get vaccinated (that they consume dietary supplements). If these proportions are substantially larger than the proportions that we directly elicit in the experiment, the self-reported attitudes might suffer from social desirability or experimenter demand bias.

Columns 1 to 4 in Table 3.12 show that the proportion of participants who directly report that they do not want to get vaccinated or boostered is similar (column 1) or even larger (column 3) than the proportion elicited through the list experiment. The proportions of participants who directly report to consume dietary supplements are smaller than the proportion that we elicit through the list experiment (columns 2 and 4), but the differences are small and could be driven by the slightly different questions in the main and in the list experiment ("How likely are you to consume dietary supplements in the near future?" vs. "I consume dietary supplements.") In sum, there is no evidence for systematic experimenter demand bias with respect to the self-reported attitudes on Corona vaccination and dietary supplements when we consider the entire sample.[67]

Next, we check for experimenter demand bias within our three main treatment groups. Given our experimental setup, experimenter demand bias is especially likely to occur for the FactChecking group when we ask about their willingness to get vaccinated in Wave I of the survey, and for the MediaLiteracy group throughout. However, as for the entire sample, the proportions of participants who directly state that they do not want to get vaccinated are even larger than the proportions of participants who indirectly say so, both for the FactChecking and for the MediaLiteracy group, and in both Wave I and Wave II of the survey. Hence, we do not find evidence for experimenter demand bias here. We do observe that for the MediaLiteracy group, the proportion of participants who just indirectly report to consume dietary supplements is larger than the proportion of participants who directly says so. However, this is also true for the FactChecking and the NoIntervention group, where experimenter demand bias is unlikely to occur, which suggests that the result is driven by unrelated issues. In sum, we can conclude that experimenter demand bias is unlikely to confound our main results.

3.6.6.4 IV analysis

Participants from the FactChecking and the MediaLiteracy group might skip the intervention by just quickly clicking through the survey. In this case, the ITT

[67]We further support the analysis with a sample split (see Table B21 in Appendix B.4), where we consider participants who directly report that they are likely to get vaccinated (consume dietary supplements) on the one hand, and participants who report that they are unlikely to get vaccinated (consume dietary supplements) on the other. Reassuringly, we find that the proportion of participants who indirectly concede that they are not going to get vaccinated against Covid-19 is much larger for the latter than for the former group. Similarly, the proportion of participants who indirectly concede to consume dietary supplements is much larger for participants who directly report that they are likely to do so than for participants who report that they are not.

Table 3.12: List experiments

	All participants				No Intervention				Fact-checking				Media Literacy			
	Wave I		**Wave II**		**Wave I**		**Wave II**		**Wave I**		**Wave II**		**Wave I**		**Wave II**	
	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.	Vacc.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Direct question	0.19	0.30	0.22	0.31	0.22	0.30	0.26	0.30	0.17	0.29	0.20	0.31	0.16	0.27	0.20	0.28
List experiment	0.16	0.38	0.08	0.33	0.21	0.34	0.11	0.31	0.14	0.37	0.07	0.45	0.06	0.40	0.01	0.26
N	3,051	3,051	2,525	2,525	618	618	509	509	607	607	513	513	613	613	511	511

Notes: In row 1, Table 3.12 displays the proportion of participants who in the main experiment **directly** report to be *Very unlikely* or *Unlikely* to get vaccinated or boostered against Covid-19 (columns 1 and 3) and who **directly** report to be *Very likely* or *Likely* to consume dietary supplements (columns 2 and 4). In row 2, Table 3.12 displays the respective indirectly elicited proportions from the list experiments.

estimates would underestimate the average treatment effect. To take this into account, this section presents an IV approach, where we use participants’ time spent with the interventions to determine their actual treatment status and their random assignment to a treatment group as an instrument.

We proceed in two steps. First, we specify how much time it takes to properly engage with the interventions. To this end, we asked eleven Research Assistants to carefully read the two fact-checks as well as the ten tips to spot false news and recorded how much time they need. We find that the minimum amount of time spent on the fact-checking intervention is equal to 24.7, and the minimum amount of time spent on the media literacy intervention is equal to 38.9 seconds.

Second, we define D_i as a dummy variable that indicates participant i’s actual treatment status. In particular, D_i is equal to one if i spent at least 24.7 seconds with the fact-checking or 38.9 seconds with the media literacy intervention. Thus, D_i is equal to zero for participants who did not spent a reasonable amount of time with their respective intervention, and for all participants in the NoIntervention group.[68] Equation (3.1) thus extends to

$$y_{iw} = \gamma_0 + \gamma_1 \widehat{D_i} + \gamma_2 X_i + \epsilon_i \tag{3.3}$$

$$D_i = \pi_0 + \pi_1 TG_i + \pi_2 X_i + u_i, \tag{3.4}$$

which we estimate by 2SLS.

We prefer using a binary (rather than a continuous) measure for participants’ actual treatment status for two reasons. First, the impact of time spent with the interventions is likely to be discrete: participants need a certain minimum amount of time to understand and process the novel information, but any time spent beyond that is unlikely to yield further benefits. Second, it generally takes more time to engage with the media literacy than with the fact-checking intervention. Hence, using a binary measure for participants’ actual treatment status makes the regression results better comparable across treatment groups.

Table B22 in Appendix B.4 confirms that the 2SLS estimates of equations (3.3) and (3.4) are larger, but qualitatively similar to their counterparts from Section 3.4.1.[69] Moreover, the coefficients for π_1 demonstrate that close to 70% of the FactChecking, and 74% of the MediaLiteracy group spent a considerable

[68] Using the minimum amount of time spent with the interventions as our threshold is the most conservative choice. When we use the median or mean amount of time from surveying the Research Assistants, the IV estimates become larger, but are qualitatively unaffected.

[69] As further robustness checks for the MediaLiteracy group, we replace D_i with a dummy that is (i) equal to one if participant i reports to have used the tips during the experiment and (ii) equal to one if participant i could recall the tips correctly. The 2SLS estimates are larger than their counterparts in Table B22 in both cases, but qualitatively unaffected.

amount of time with their respective intervention. Skipping the interventions could be a larger concern outside the context of our experiment, though, especially if they disrupt users' consumption of social media. We further discuss this issue below.

3.7 Conclusion

We conduct a large-scale randomized survey experiment on the immediate and short-term effects of fact-checking and media literacy interventions to demonstrate that the impact of fact-checking tends to be limited to the fakes that are being corrected, whereas the media literacy intervention helps to distinguish between fakes and facts more generally, both immediately and in the short-run. A plausible mechanism for our result is that the media literacy intervention enables participants to critically evaluate social media postings, while fact-checking turns them into passive recipients of the specific corrections and thus fails to markedly enhance their skills. Hence, in an environment where not every claim can be fact-checked, the media literacy intervention is likely to be more effective than fact-checking on average.

Our paper promotes brief media literacy interventions as an effective tool to fight fake news and advances current policy debates along these lines. The European Union, for instance, has recently asserted media literacy as a pivotal tool to counter misinformation on social media,[70] and the UNESCO has provided policy guidelines for digital media and information literacy.[71] Our results strongly support such endeavors and suggest that official media literacy campaigns – which are relatively cheap, scalable, and easy-to-implement – could be a valuable complement to existing efforts like fact-checking.

Media literacy interventions as a means to fight fake news might outperform fact-checking for three more reasons. First, many news items are not clearly fake or fact. While fact-checkers can only debunk fake news that clearly provide wrong or misleading information, media literacy interventions are likely to raise users' skills and awareness to spot even more subtle types of misinformation. Second, the effectiveness of fact-checking hinges on users' trust in the fact-checker. E.g., about 50% of Americans think that fact-checkers are biased (Allen et al., 2021). In times where users' trust in renowned sources like public service broadcasters is crumbling, the provision of fact-checks is likely to be futile, as those whom the intervention is supposed to target are often those who distrust the fact-checker. Third and relatedly, with the rise of Artificial Intelligence (AI), users might have to increasingly

[70] See `https://digital-strategy.ec.europa.eu/en/policies/media-literacy` (viewed August 2022).

[71] `https://www.unesco.org/en/communication-information/media-information-literacy/policy-strategy` (viewed August 2022).

rely on their own critical judgment of what they see rather than blindly trust the assessment of third parties. This applies even when AI is being used by researchers and journalists to *assist* the detection of fake news. Specifically, while AI is currently able to detect suspicious and check-worthy news items and match them to claims that have previously been fact-checked, fact-checking as such is still based on manual assessment and far from being fully automated.[72]

Our analysis has several limitations that open avenues for further research. First, the magnitude of our coefficients is likely to depend on the specific fakes, facts, and fact-checks as well as on the topics that we selected for the experiment, and other details of the implementation such as the absence of placebo treatments. Therefore, we consider the qualitative results as our most insightful findings and recommend to interpret the precise point estimates with caution. E.g., some fakes are harder to detect than others, which is likely to reduce the effectiveness of our interventions. Also, there exist different types of fact-checking (e.g., journalist-based vs. crowd-sourced fact-checking, see Allen et al. (2021)), whose effectiveness might vary in different contexts and depend on the recipient. Similarly, users may be more or less well informed about different topics, whereby they are more or less likely to benefit from our interventions.

Our sample is representative of the German population in terms of age and gender, suggesting that our findings are generalizable to countries with similar demographics, such as many in the EU [73]. Notably, surveys suggest no heightened sensitivity among German users to fake news or media literacy initiatives. For example, evidence from *RavenPack* indicates that German news media discuss fake news related to Covid-19 less frequently than other EU countries.[74] Furthermore, Germans report encountering fake news less often than respondents from other countries.[75] While fake news about dietary supplements is prevalent on social media[76], 56% of Germans feel well-informed about the associated risks and benefits.[77] Additionally, although Germany implemented the Network Enforcement Act in 2017, its impact on user behavior, particularly on Facebook, appears minimal (Maaß et al., 2024). While we cannot rule out that our results are specific for a German audience, the mentioned evidence lowers our concerns about a limited external validity beyond the German

[72]See EU Horizon, URL: `https://ec.europa.eu/research-and-innovation/en/horizon-magazine/can-artificial-intelligence-help-end-fake-news` (viewed August 2023).

[73]See Eurostat: Demography of Europe – 2023 (viewed December 2024).

[74]See `https://coronavirus.ravenpack.com/` (viewed April 2024).

[75]See survey results by *Statista*: 1 , 2 (viewed December 2024).

[76]See media articles by Eisele (Deutsche Welle, viewed December 2024: *Faktencheck: Abzocke mit Nahrungsergänzungsmitteln?* and Schaetze and Lerch (consumer protection agency, viewed December 2024: *Gesundheitsversprechen für Nahrungsergänzungsmittel auf Instagram – häufig abseits der Legalität*).

[77]See *Forsa Survey on Opinions on Nutrition Supplements* (viewed December 2024).

context – especially regarding the qualitative results.

Second, it is unclear how many and which users would actually engage with media literacy interventions. In particular, some users might perceive such trainings as a nuisance and consequently skip them. In addition, it could be that mostly users who are well informed anyway decide to take part in media literacy interventions, while users with poor priors – i.e., those for whom the intervention would be most effective – prefer to shirk them. Participation in media literacy interventions will ultimately depend on their design. However, even if such interventions fail to reach the entire population, it is worthwhile to enhance the skills even of a subset of users, and this should be preferred over not doing anything. Thus, while our paper stresses the high potential of media literacy interventions as a tool to fight fake news, the most efficient ways to implement such interventions must be examined in future research.

Third and relatedly, while we demonstrate that media literacy interventions could help users to better distinguish between false and correct information that they encounter online, we remain agnostic about the concrete implementation of such interventions on behalf of social media platforms. In particular, it is unclear if social media would be willing to set up regular interventions (e.g., in terms of pop-up windows that appear every few weeks) and what the ideal type of intervention would look like. The fact that Facebook has developed a set of "Tips to Spot False News" on its own behalf is encouraging, though, and suggests that social media might be willing to cooperate with academics and policy makers. The ideal type of intervention is likely to depend on the specific social media platform – e.g., users on TikTok may require different tips than users on Facebook – and promises to be an interesting field for future research.

Data Statement

The experiment was approved by the Ethics Committee of the University of Cologne (reference 210014JM) and pre-registered in the AEA Registry under registry number AEARCTR-0008199. It was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1-390838866 and by the Bavarian State Ministry of Science and the Arts in the framework of the bidt Graduate Center for Postdocs. A replication package including all anonymized datasets and code is deposited at https://doi.org/10.17605/OSF.IO/SGMJN.

Bibliography for Chapter 3

Alesina, A., A. Miano, and S. Stantcheva (2023). “Immigration and Redistribution”. In: *The Review of Economic Studies* 90.1, pp. 1–39.

Ali, S., M. H. Saeed, E. Aldreabi, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini (2021). “Understanding the effect of deplatforming on social networks”. In: *Proceedings of the 13th ACM Web Science Conference 2021*, pp. 187–195.

Allcott, H. and M. Gentzkow (2017). “Social Media and Fake News in the 2016 Election”. In: *Journal of Economic Perspectives* 31.2, pp. 211–236.

Allen, J., D. J. Watts, and D. G. Rand (2024). “Quantifying the impact of misinformation and vaccine-skeptical content on Facebook”. In: *Science* 384.6699, pp. 1–8.

Allen, J., A. A. Arechar, G. Pennycook, and D. G. Rand (2021). “Scaling up fact-checking using the wisdom of crowds”. In: *Science Advances* 7.36, pp. 1–10.

Bailey, R. L., J. J. Gahche, P. E. Miller, P. R. Thomas, and J. T. Dwyer (2013). “Why US adults use dietary supplements”. In: *JAMA Internal Medicine* 173.5, pp. 355–361.

Bak-Coleman, J. B., I. Kennedy, M. Wack, A. Beers, J. S. Schafer, E. S. Spiro, K. Starbird, and J. D. West (2022). “Combining interventions to reduce the spread of viral misinformation”. In: *Nature Human Behaviour* 6.10, pp. 1372–1380.

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). “Facts, alternative facts, and fact checking in times of post-truth politics”. In: *Journal of Public Economics* 182, p. 104123.

Blair, G. and K. Imai (2012). “Statistical analysis of list experiments”. In: *Political Analysis* 20.1, pp. 47–77.

Bode, L. and E. K. Vraga (2015). “In related news, that was wrong: The correction of misinformation through related stories functionality in social media”. In: *Journal of Communication* 65.4, pp. 619–638.

Brashier, N. M., G. Pennycook, A. J. Berinsky, and D. G. Rand (2021). “Timing matters when correcting fake news”. In: *Proceedings of the National Academy of Sciences* 118.5.

Broniatowski, D. A., J. R. Simons, J. Gu, A. M. Jamison, and L. C. Abroms (2023). “The efficacy of Facebook’s vaccine misinformation policies and architecture during the COVID-19 pandemic”. In: *Science Advances* 9.37.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2023). “Opinions as Facts”. In: *Review of Economic Studies* 90.4, pp. 1832–1864.

Carey, J. M., A. M. Guess, P. J. Loewen, E. Merkley, B. Nyhan, J. B. Phillips, and J. Reifler (2022). “The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada". In: *Nature Human Behaviour* 6.2, pp. 236–243.

Cheung, A. C. and R. E. Slavin (2016). "How methodological features affect effect sizes in education". In: *Educational Researcher* 45.5, pp. 283–292.

Chiou, W.-B., C.-C. Yang, and C.-S. Wan (2011). "Ironic effects of dietary supplementation: illusory invulnerability created by taking dietary supplements licenses health-risk behaviors". In: *Psychological Science* 22.8, pp. 1081–1086.

Clarke, D., J. P. Romano, and M. Wolf (2020). "The Romano–Wolf multiple-hypothesis correction in Stata". In: *The Stata Journal* 20.4, pp. 812–843.

Dai, Y., W. Yu, and F. Shen (2021). "The effects of message order and debiasing information in misinformation correction". In: *International Journal of Communication* 15, pp. 1039–1059.

De Quidt, J., J. Haushofer, and C. Roth (2018). "Measuring and bounding experimenter demand". In: *American Economic Review* 108.11, pp. 3266–3302.

DellaVigna, S. and M. Gentzkow (2010). "Persuasion: Empirical Evidence". In: *Annual Review of Economics* 2.1, pp. 643–669.

DellaVigna, S. and E. Kaplan (2007). "The Fox News effect: Media bias and voting". In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

Deslauriers, L., E. Schelew, and C. Wieman (2011). "Improved learning in a large-enrollment physics class". In: *Science* 332.6031, pp. 862–864.

Drexler, A., G. Fischer, and A. Schoar (2014). "Keeping it simple: Financial literacy and rules of thumb". In: *American Economic Journal: Applied Economics* 6.2, pp. 1–31.

Drolsbach, C., K. Solovev, and N. Pröllochs (2024). "Community notes increase trust in fact-checking on social media". In: *PNAS Nexus* 3.7, pgae217.

Ecker, U. K., S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen (2022). "The psychological drivers of misinformation belief and its resistance to correction". In: *Nature Reviews Psychology* 1.1, pp. 13–29.

Erbaugh, J. T., C. H. Chang, Y. J. Masuda, and J. Ribot (2024). "Communication and Deliberation for Environmental Governance". In: *Annual Review of Environment and Resources* 49, pp. 367–393.

Ershov, D. and J. S. Morales (2024). "Sharing News Left and Right: Frictions and Misinformation on Twitter". In: *The Economic Journal* 134.662, pp. 2391–2417.

Fong, J., T. Guo, and A. Rao (2024). "Debunking Misinformation About Consumer Products: Effects on Beliefs and Purchase Behavior". In: *Journal of Marketing Research* 61.4, pp. 659–681.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth (2014). “Active learning increases student performance in science, engineering, and mathematics”. In: *Proceedings of the National Academy of Sciences* 111.23, pp. 8410–8415.

Fryer Jr, R. G. (2017). “The production of human capital in developed countries: Evidence from 196 randomized field experiments”. In: *Handbook of economic field experiments*. Vol. 2. Elsevier, pp. 95–322.

Gensing, P. (2021). “Wie die AfD Angst vor Impfungen schürt”. In: *Tagesschau.de*.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). “Fake news on Twitter during the 2016 US presidential election”. In: *Science* 363.6425, pp. 374–378.

Gu, J., A. Dor, K. Li, D. A. Broniatowski, M. Hatheway, L. Fritz, and L. C. Abroms (2022). “The impact of Facebook’s vaccine misinformation policy on user endorsements of vaccine content: An interrupted time series analysis”. In: *Vaccine* 40.14, pp. 2209–2214.

Guess, A., J. Nagler, and J. Tucker (2019). “Less than you think: Prevalence and predictors of fake news dissemination on Facebook”. In: *Science Advances* 5.1, eaau4586.

Guess, A., B. Nyhan, and J. Reifler (2018). “Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign”. In: *European Research Council* 9.3, p. 4.

Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). “A digital media literacy intervention increases discernment between mainstream and false news in the United States and India”. In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15536–15545.

Gundersen, T., D. Alinejad, T. Y. Branch, B. Duffy, K. Hewlett, C. Holst, S. Owens, F. Panizza, S. M. Tellman, J. van Dijck, and M. Baghramian (2022). “A new dark age? Truth, trust, and environmental science”. In: *Annual Review of Environment and Resources* 47.1, pp. 5–29.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). “Curtailing False News, Amplifying Truth”. In: *Working Paper*.

Henry, E., E. Zhuravskaya, and S. Guriev (2022). “Checking and Sharing Alt-Facts”. In: *American Economic Journal: Economic Policy* 14.3, pp. 55–86.

Hill, C. J., H. S. Bloom, A. R. Black, and M. W. Lipsey (2008). “Empirical benchmarks for interpreting effect sizes in research”. In: *Child development perspectives* 2.3, pp. 172–177.

Hopkins, D. J. (2009). “No more Wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates”. In: *The Journal of Politics* 71.3, pp. 769–781.

Jerit, J. and Y. Zhao (2020). “Political misinformation”. In: *Annual Review of Political Science* 23, pp. 77–94.

Kaiser, T., A. Lusardi, L. Menkhoff, and C. Urban (2022). “Financial education affects financial knowledge and downstream behaviors”. In: *Journal of Financial Economics* 145.2, pp. 255–272.

Kaiser, T. and L. Menkhoff (2022). “Active learning improves financial education: Experimental evidence from Uganda”. In: *Journal of Development Economics* 157.

Karlin, B., J. F. Zinger, and R. Ford (2015). “The effects of feedback on energy conservation: A meta-analysis.” In: *Psychological Bulletin* 141.6, pp. 1205–1227.

Kuklinski, J. H., P. J. Quirk, J. Jerit, D. Schwieder, and R. F. Rich (2000). “Misinformation and the currency of democratic citizenship”. In: *The Journal of Politics* 62.3, pp. 790–816.

Lazer, D. M., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, and D. Rothschild (2018). “The science of fake news”. In: *Science* 359.6380, pp. 1094–1096.

Lewandowsky, S. (2021). “Climate change disinformation and how to combat it”. In: *Annual Review of Public Health* 42.1, pp. 1–21.

Lewandowsky, S., U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook (2012). “Misinformation and its correction: Continued influence and successful debiasing”. In: *Psychological science in the public interest* 13.3, pp. 106–131.

Lewandowsky, S. and S. Van Der Linden (2021). “Countering misinformation and fake news through inoculation and prebunking”. In: *European Review of Social Psychology* 32.2, pp. 348–384.

Liu, X., L. Qi, L. Wang, and M. J. Metzger (2023). “Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness”. In: *Communication Research*.

Loewenstein, G. and Z. Wojtowicz (2023). “The Economics of Attention”. In: *Available at SSRN 4368304*.

Luca, M. (2015). “User-generated content and social media”. In: *Handbook of Media Economics*. Vol. 1. Elsevier, pp. 563–592.

Maaß, S., J. Wortelker, and A. Rott (2024). “Evaluating the regulation of social media: An empirical study of the German NetzDG and Facebook”. In: *Telecommunications Policy* 48.5, p. 102719.

Maertens, R., J. Roozenbeek, M. Basol, and S. van der Linden (2021). "Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments." In: *Journal of Experimental Psychology: Applied* 27.1, p. 1.

Martel, C. and D. G. Rand (2024). "Fact-checker warning labels are effective even for those who distrust fact-checkers". In: *Nature Human Behaviour* 8, pp. 1957–1967.

Mitts, T., N. Pisharody, and J. Shapiro (2022). "Removal of anti-vaccine content impacts social media discourse". In: *Proceedings of the 14th ACM Web Science Conference 2022*, pp. 319–326.

Mummolo, J. and E. Peterson (2019). "Demand effects in survey experiments: An empirical assessment". In: *American Political Science Review* 113.2, pp. 517–529.

Noar, S. M., C. N. Benac, and M. S. Harris (2007). "Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions." In: *Psychological Bulletin* 133.4, p. 673.

Nyhan, B. (2021). "Why the backfire effect does not explain the durability of political misperceptions". In: *Proceedings of the National Academy of Sciences* 118.15, e1912440117.

Nyhan, B., E. Porter, J. Reifler, and T. J. Wood (2020). "Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability". In: *Political Behavior* 42.3, pp. 939–960.

Nyhan, B. and J. Reifler (2010). "When corrections fail: The persistence of political misperceptions". In: *Political Behavior* 32.2, pp. 303–330.

— (2015). "Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information". In: *Vaccine* 33.3, pp. 459–464.

Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2021). "Shifting attention to accuracy can reduce misinformation online". In: *Nature* 592.7855, pp. 590–595.

Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand (2020). "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention". In: *Psychological Science* 31.7, pp. 770–780.

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

— (2021). "The Psychology of Fake News". In: *Trends in Cognitive Sciences* 25.5, pp. 388–402.

Radimer, K., B. Bindewald, J. Hughes, B. Ervin, C. Swanson, and M. F. Picciano (2004). "Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000". In: *American Journal of Epidemiology* 160.4, pp. 339–349.

Romano, J. P. and M. Wolf (2005). "Stepwise multiple testing as formalized data snooping". In: *Econometrica* 73.4, pp. 1237–1282.

Rooney, B. L. and D. M. Murray (1996). "A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis". In: *Health education quarterly* 23.1, pp. 48–64.

Roozenbeek, J., S. van der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky (2022). "Psychological inoculation improves resilience against misinformation on social media". In: *Science Advances* 8.34, eabo6254.

Ruiz-Primo, M. A., D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard (2011). "Impact of undergraduate science course innovations on learning". In: *Science* 331.6022, pp. 1269–1270.

Stantcheva, S. (2021). "Understanding tax policy: How do people reason?" In: *The Quarterly Journal of Economics* 136.4, pp. 2309–2369.

Swire, B., A. J. Berinsky, S. Lewandowsky, and U. K. Ecker (2017). "Processing political misinformation: comprehending the Trump phenomenon". In: *Royal Society Open Science* 4.3.

Swire-Thompson, B. and D. Lazer (2020). "Public health and online misinformation: challenges and recommendations". In: *Annual Review of Public Health* 41.1, pp. 433–451.

Swire-Thompson, B., J. Cook, L. H. Butler, J. A. Sanderson, S. Lewandowsky, and U. K. Ecker (2021). "Correction format has a limited role when debunking misinformation". In: *Cognitive Research: Principles and Implications* 6.83, pp. 1–15.

Tversky, A. and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty." In: *Science* 185.4157, pp. 1124–1131.

Verplanken, B. and S. Orbell (2022). "Attitudes, habits, and behavior change." In: *Annual Review of Psychology* 73, pp. 327–352.

Vosoughi, S., D. Roy, and S. Aral (2018). "The spread of true and false news online". In: *Science* 359.6380, pp. 1146–1151.

Vraga, E. K. and L. Bode (2017). "Using expert sources to correct health misinformation in social media". In: *Science Communication* 39.5, pp. 621–645.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). "Political effects of the internet and social media". In: *Annual Review of Economics* 12, pp. 415–438.

Chapter 4

Improving science literacy in the newsroom: Experimental evidence

Paper Information

Authors: Lara Marie Berger, Dr. Anna Kerkhof, Nikola Noske

Status: Published in PNAS Nexus, Volume 5, Issue 4, April 2026, pgag099.

Abstract

Accurate reporting of new findings is essential for an informed public, yet limited science literacy among journalists often results in misinterpretation of research, with costly societal and economic consequences. We developed a concise educational video that aimed to enhance journalists' science literacy through explaining important checkpoints for sound research reporting (funding, sample, statistics, causal claims, visuals). The impact of this video was tested in a survey experiment with 260 German journalists. Treated participants produced correct headlines for 64 percent of study-based news stories versus 36 percent in the control group (a 28-percentage-point rise, $p < 0.001$). They also became more skeptical when evaluating existing articles, flagging more potential mistakes. An exploratory follow-up using 268 real-world articles suggests a reduction in factual mistakes by treated journalists. Our findings highlight that a brief, low-cost intervention can measurably improve the scientific accuracy of journalism, offering a scalable strategy for news organizations and training programs.

4.1 Introduction

Scientific findings can substantially improve public welfare. For example, medical innovations often help to reduce or prevent diseases. Data can enhance the effectiveness of public policy, optimize resource allocation, and inform economic decisions in the interest of the public good. However, decision makers usually don't rely directly on academic publications. The often technical and specialized language poses a barrier, and closed-access publications, as well as researchers' limited resources for science communication further restrict public access to new discoveries. In this context, journalists play a pivotal role: They act as translators and multipliers of scientific knowledge, by transforming hard facts into more engaging narratives. Thereby, journalists have the potential to bridge the gap between science and society.

Yet, journalists' crucial mediating role comes with significant challenges that undermine the accurate transmission of scientific insights. A prominent example illustrates the stakes: The "Mozart effect". Rauscher et al. (1993) cautiously concluded that listening to Mozart can temporarily enhance spatial reasoning, but that this effect does not extend beyond the 10-15-minute period during which subjects were engaged in a spatial task in their experiment. Media reports soon exaggerated this finding: A widely circulated *New York Times* article claimed listening to Mozart "makes you smarter", without mentioning the short time span and other limitations. In turn, the governor of Georgia demanded to distribute classical music CDs to all mothers of newborns – at a cost of more than 100.000 dollars. This cascade from a nuanced result to public misperception underscores how even well-intentioned reporting can distort scientific meaning and lead to misguided decisions. While this example may seem relatively harmless in retrospect, the consequences of miscommunication are far more serious in domains such as climate change or vaccination campaigns, where public misunderstanding of science can have severe lasting societal effects.

Dahlstrom (2021) thus emphasizes the existence of a trade-off: While storytelling enhances the visibility and spread of scientific information, it may also foster misconceptions by inaccurate translations of facts. Indeed, miscommunication of science in news reports has been found to be widespread (Oxman et al., 2022) and amplified by systematic issues in media markets. Digitization rises economic pressures and high resource constraints mean that many journalists lack specialized training in science and statistics, work under tight deadlines and heavily rely on press releases: a practice that fosters sensationalism and inaccuracies (Armona et al., 2024; Dempster et al., 2022; Serra-Garcia, 2025; Sumner et al., 2016).

How can the accuracy in journalistic reporting be improved? Addressing this gap, we partnered with a journalism school to design a short educational video with the aim to enhance science literacy among journalists. The video encouraged journalists to critically think about a studies (i) funding, (ii) sample composition, (iii) statistics, (iv) causal interpretations and (v) illustrations and graphs. The effectiveness of watching this video was tested in a randomized, pre-registered online survey experiment involving $N = 260$ professional journalists from Germany.

In this experiment we randomly displayed our video to half of the journalists before proceeding. For ethical reasons the control-group received a link to the video after taking part in the survey. Subsequently, we exposed participants to scientific studies which have previously been found to be especially prone to misinterpretation. We also exposed participants to real media articles which have been published about those studies. Then, we asked them to write a new headline for each given news article and to evaluate the accuracy of the article regarding the interpretation of the original study in different dimensions. Afterwards, we assessed the accuracy of the written headlines by having research assistants determine whether a headline contains any factual mistake. We also compared the journalists' article evaluations with those of our research assistants.

We found that the overall accuracy of produced headlines in our sample was low. Only 36 percent of headlines in the control group were factually correct, highlighting journalists' difficulties in interpreting the scientific studies. However, journalists who watched our seven-minute educational video were 28 percentage points more likely to write an accurate headline than those in the control group ($p < 0.001$), raising the share of correct headlines in the treatment group to 64 percent on average. The size of this effect corresponds to 0.6 standard deviations in the control group. One plausible explanation for the comparatively large effect size is that the journalists examined studies especially susceptible to false interpretations, leaving ample room for our intervention to improve reporting accuracy.

Furthermore, journalists were inherently skeptical of the reports they evaluated: In the control group, 54 percent expected an article to contain at least one factual mistake. Our intervention amplified this skepticism. Treated journalists were 10 percentage points more likely to classify an article as at least partially incorrect ($p = 0.019$, 0.22 standard deviations in the control group) and expect, on average, 0.45 more mistakes per article ($p = 0.001$, 0.32 standard deviations in the control group). This pattern indicates that the intervention fostered a more generalized skepticism rather than clear improvements in discernment relative to our pre-registered benchmarks. In other words, while treated journalists were more inclined to flag potential errors, we did not find consistent evidence that they became better at distinguishing between accurate and inaccurate reporting. We return to this distinction in more detail in Section 4.5. While we did not find evidence that watching the video improves accuracy in identifying the exact number of errors, our data supports an enhancement in the ability to classify certain error types. Specifically, treated journalists were 20 percentage points ($p = 0.016$) more likely to identify statistical misinterpretations and 10 percentage points ($p = 0.133$) more likely to recognize confusions between correlation and causation. We constructed a standardized Z-Score index to aggregate all our pre-registered outcomes (the accuracy of article classification based on pre-registered benchmarks, the absolute deviation from the pre-registered number of mistakes per article, the correct classification of error types, and headline accuracy). Being in the treatment group improved this index by 0.12 standard deviations ($p = 0.029$).

To assess a potential real-world impact of our intervention, we tracked the published work of the 64 journalists who consented to share their names and employers for the purpose of our analysis. We compiled a database of up to ten articles they publish in the two months following their participation in our survey and analyze whether (i) the article focused on a scientific study and (ii) it cited at least one scientific study to support a claim. The results highlight the marginal role of scientific research in journalistic reporting: 84 percent of the articles did not cover science as a main topic, and 67 percent did not cite any scientific study. This is in line with studies documenting that science reporting tends to make up only a small fraction of overall news reports, e.g. Suleski and Ibaraki (2010),Vestergaard and Nielsen (2016) or Summ and Volpers (2016).

For the small subset of articles that either are about a study or cite a scientific study at some point (a total of 58 articles written by 21 journalists), we use *ChatGPT*'s deep research function to classify whether the reporting was accurate. This analysis revealed that most of the real articles written by these journalists were correct: Only 14 percent contained at least one mistake. As some of the control-group participants accessed our video after the survey, we estimated the effect of our intervention on the likelihood of a real-world article to contain at least one mistake with an instrumental variable approach. This estimation suggests that our intervention reduced the likelihood of a real-world article to contain at least one mistake ($p = 0.020$). Due to the small and likely very selective sample of journalists and articles in this analysis we recommend interpreting this exploratory finding cautiously.

Overall, our findings highlight systemic challenges in journalistic science reporting. Even trained professionals struggled to write accurate headlines and critically assess scientific claims, underscoring the economic and societal risks of misinterpreted science. However, our results also demonstrated that a brief, low-cost intervention significantly improved headline accuracy on average. This aligns with existing evidence on the effectiveness of brief, targeted approaches to prevent the spread of misinformation (L. M. Berger et al., 2025; A. M. Guess et al., 2020; Markowitz, 2024; Roozenbeek et al., 2022) and extends prior research by showing that such interventions can be effectively tailored to science journalism. Beyond academic contributions, our study has practical implications for professional training in journalism: News organizations, journalism schools, and professional associations could integrate similar short, scalable training modules into curricula and professional development programs. Given the role of journalists in shaping public discourse and policy debates, improving their ability to critically evaluate scientific evidence is essential to counter misinformation and foster a better understanding of science.

The remainder of the paper is structured as follows: Section 4.2 reviews related literature. Section 4.3 details the experimental setup, implementation, and empirical strategy. Section 4.4 presents our findings, including robustness checks. Section 4.5 discusses the effect sizes, implications and limitations of our findings. Section 4.6 concludes.

4.2 Related Literature

Our paper contributes to two strands of literature. The first discusses the importance of and challenges with science journalism for an enlightened society. Dahlstrom (2021) emphasizes the existence of a fundamental trade-off: While storytelling enhances the visibility and reach of scientific findings, it also fosters misconceptions by (most likely unintended) inaccurate translations. This is problematic, as news coverage has been found to impact several outcomes relevant for the public good such as economic fluctuations, political decisions or health related behaviors, see e.g. Chahrour et al. (2021), **Bursztynetal2023**, Durante and Zhuravskaya (2018) or Eisensee and Strömberg (2007). Cacciatore (2021) highlights that misinformation in science related domains is consistently interpreted as a major concern and that it is considered challenging to eliminate.

Journalists are key intermediaries in translating scientific findings into accessible narratives for the public. However, the literature suggests that systematic challenges like economic pressures, tight deadlines and reliance on press releases often hinder the accurate representation of research in news reports. For example, economic incentives to generate attention have dramatically risen due to digitization of news markets (Wu, 2016) and Serra-Garcia (2025) shows that such incentives reduce the amount of accurate information in reports about science. Sumner et al. (2016) emphasize the role of mistakes in press-releases and show that they often translate into mistakes in news reports. Also, robust, but less sensational stories have been found to be less newsworthy (Armona et al., 2024), whereas early-stage or controversial research tends to dominate headlines (Selvaraj et al., 2014). Furthermore, misinterpretations of statistical findings, such as confusing correlation with causation, are common. For example, Dempster et al. (2022) analyze the coverage of a scientific study and report that nearly half of the related articles misinterpret the findings they cover. Oxman et al. (2022) provide a meta-analysis regarding the spread of mistakes in science reporting in health related contexts and conclude that overall misrepresentation of science in news is widespread.

While this literature stresses the existing risks of the transmission of inaccuracies when news reporters write about science, little rigorous work exists on what can be done to mitigate those risks. Our paper contributes to filling this gap by providing causal evidence for one promising strategy: the enhancement of science literacy of multiplicators of knowledge – professional journalists. As key intermediaries between research and the public, better-equipped journalists could serve as a supply-side antidote to misinformation, raising the quality of science reporting despite the pressures of modern media markets.

The second literature strand we contribute to is on brief educational interventions as means to mitigate the harm of false information. There is strong causal evidence showing that short educational interventions can effectively mitigate the harm of fake news. A. M. Guess et al. (2020) and L. M. Berger et al. (2025) for example show that brief tips to spot fake news on social media reduce their persuasiveness. Roozenbeek et al. (2022) find similar effects for short video interventions on the sharing rates of misinformation. Guriev et al., 2023 show that an attention nudge can reduce sharing of fake content on X (formerly Twitter). Markowitz, 2024 finds that short, AI-generated summaries of scientific research reduce errors layman make when citing these studies.

Most of these studies focus on social media environments and layman as content producers. Little is known on whether these remedies are also effective in the context of the production of professional news. This question is of great interest, because journalists, while having a much higher reach than the average social media user, have been found to show attitudes and behaviors which are not at all representative for the general population. For example, journalists are typically highly educated and hold more liberal world views than the average citizen (Hanitzsch et al., 2016; Kirkegaard et al., 2021; Willnat et al., 2022). It is therefore unclear if and how the rather short and simplistic interventions typically tested in the literature also work among this population. Also, experimental studies involving professional journalists as participants are scarce to begin with (exceptions are Balbuzanov et al. (2025), Serra-Garcia (2025) and Graves et al. (2016)), even though they are much more likely to influence public policy with their content than laymen. We thus extend the literature on educational interventions as a means again misinformation by taking a targeted educational intervention to the highly relevant, yet under-researched context of professional journalism.

4.3 Experimental Design and Data

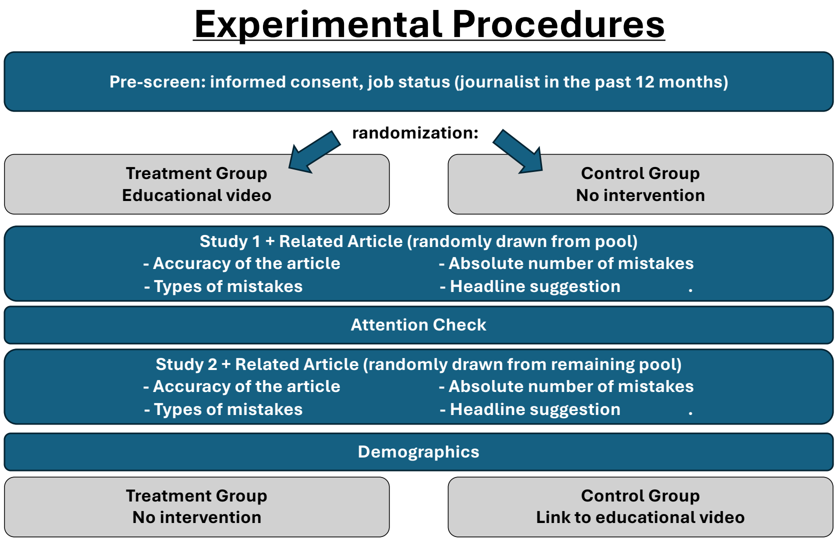
We recruited participants via an email sent to the mailing list of the *German Journalistic Association*, a professional organization that admits only individuals who can verify their status as working journalists. Our invitation informed recipients that they would have to complete a survey and receive 10 € for participating. Furthermore, they were told that they would receive information on science reporting through their participation. We did not restrict the sample to journalists with formal editorial responsibility for headlines, as this would have substantially reduced sample size and power given the relatively small share of journalists in exclusive editor roles. The recruitment email was distributed on August 13, 2024. As pre-registered, we planned to accept responses until either 300 participants had completed the survey or August 31, 2024, whichever came first. By the end of August, a total of $N = 260$ journalists had completed the study. Upon enrollment, participants were randomly assigned to either the treatment group ($N = 125$) or the control group ($N = 135$). The survey was implemented using the *Qualtrics* platform, and payments were processed via PayPal.

4.3.1 Experimental Design

Figure 4.1 outlines the procedures of the survey experiment. Participants were eligible to take part only if they reported having worked as journalists within the past 12 months and provided informed consent. Following this pre-screening, participants in the treatment group were shown the educational video. For these participants, the button to proceed with the survey appeared after three minutes, although they were free to spend additional time viewing the video. Further details on the video content are provided in section 4.3.1.1.

Next, participants received access to scientific studies that have been shown to be particularly susceptible to misinterpretation, along with real media articles reporting on these studies. To construct our pool of studies and articles, we drew on materials featured in the *Unstatistik des Monats*, a publication by the German research institute *RWI – Leibniz Institute for Economic Research* that highlights statistical misrepresentations in media reporting. This source was selected because

Figure 4.1: Experimental Procedures - Overview



Notes: Figure 4.1 provides an overview of the experimental procedures, with arrows indicating random assignment to the treatment and control groups. Scientific studies and corresponding news articles were randomly drawn from a pool of 14 reports covering five distinct studies. Each participant received access to two different studies and related articles. Participants who failed to pass a simple attention check were excluded from further participation.

our design required media articles that include at least some factual inaccuracies in their presentation of scientific findings. Further details on the selection and composition of the article pool are provided in Section 4.3.1.2.

During the survey, each participant was exposed to two different studies and their corresponding media articles. The articles were randomly selected from a pool of 14 genuine news reports published in German media, covering five distinct studies. To limit survey duration, each journalist engaged with only two studies. This design, which draws on a diverse set of articles and studies, enhances external validity of our findings. While participants had the option to access the full study via link, the information necessary to evaluate the correctness of the article was contained in the provided study excerpt. After reviewing the studies and articles, participants were asked to write a new headline for each article and to evaluate the accuracy of the article's interpretation of the original study across several dimensions.

At the end of the survey, we collected demographic information and, on a voluntary basis, invited participants to provide their name, the media outlet they write for, and consent to having their published work analyzed for a period of two months following participation. For ethical reasons, participants in the control group were offered access to the educational video via a link at the conclusion of the survey. Only 37 percent of control group participants chose to click on the link. A translation of all experimental instructions is provided in Appendix C.8.

4.3.1.1 Treatment: Video

The core of our experiment is a seven-minute educational video shown to participants in the treatment group, designed to promote critical engagement with scientific articles and statistical information.

The educational video offers a brief, practical guide for journalists on how to critically assess scientific studies and the way they are reported. It introduces six core questions that cover funding sources, sample representativeness, the interpretation of metrics, use of absolute versus relative figures, the distinction between correlation and causation, and the design of graphs. These criteria reflect established principles of science journalism and focus on common sources of misinterpretation in applied research reporting, where findings are often implicitly presented as broadly generalizable. Through simple examples and visual illustrations, the video demonstrates how misleading interpretations can arise even from accurate numbers. The video emphasizes that statistical terms like “average” can obscure important nuances such as inequality, and that visual elements like graphs or percentages can be intentionally or unintentionally deceptive. It concludes by encouraging viewers to apply these rules to improve the accuracy and clarity of their own reporting.

The content was developed based on components identified during a pre-study in cooperation with a Journalism School (see Appendix C.3 for a detailed description) and was produced with the support of the communications department at the *ifo Institute* (Leibniz Institute for Economic Research at the University of Munich). An English translation of the full video script is provided in Appendix C.5 The original video (in German) is available online: `https://youtu.be/0R322UjAY9Q`.

Participants in the control group did not have access to the video during the main study phase but were given the option to view it after completing the survey by clicking on a link. Importantly, they were not actively prompted to do so, and we recorded which participants chose to access the video. This design ensured equal access to the training material while preserving the ability to identify a (self-selected) subset of participants who remained unexposed, enabling further analyses (see Section 4.4.2.5).

4.3.1.2 Pool of Studies and Articles

Participants evaluated two news articles reporting on scientific studies, randomly drawn from a pool of 14 pre-selected articles. These articles cover the results of five distinct studies and are paired with the corresponding original research.

To construct this article pool, research assistants compiled a database of recent media reports on scientific findings, focusing on studies featured in the *Unstatistik des Monats* (English: non-statistic of the month), a publication by the *RWI – Leibniz Institute for Economic Research* that highlights statistical misinterpretations in the media. Articles were eligible if they were published within the past year and directly referenced a scientific study. We further restricted the pool to studies for which multiple news reports were available and excluded those on emotionally charged topics, in order to minimize the risk that participants' evaluations would be influenced by personal values rather than attention to statistical accuracy.

This procedure yielded a final set of 14 news articles covering five studies. It is important to note that these studies are unlikely to be representative of the broader scientific literature; any observed effects of the intervention should be interpreted in light of this targeted selection. English translations of the articles, along with links to the original media reports and studies, are provided in Appendix C.6.

As a means to keep the duration of the survey in a reasonable time frame, only the most relevant sections such as the title, teaser, and introduction were directly displayed within the survey. Full texts were accessible via dropdown menus, and we recorded whether participants chose to view the complete content.

All articles were pre-assessed by the research team for reporting errors, categorized as (i) misinterpretation of statistics, (ii) confusion between correlation and causation, (iii) lack of contextual information, and (iv) issues related to non-representative samples or selection bias. These categories, along with a classification indicating whether each article contained any clear error, were pre-registered prior to data collection.

This setup implies that our pool of materials contained both more and less accurate reporting, allowing us to construct outcomes that capture not only whether participants were skeptical in general, but also whether their assessments matched the actual accuracy of the articles. In other words, our design makes it possible to approximate discernment-like measures.

4.3.1.3 Pre-registered Outcomes

We pre-registered a set of four outcomes: (i) the detection of the presence of errors, (ii) the detection of the total number of errors identified, (iii) the detection of present error types and (iv) the accuracy of the newly created headlines. In our pre-registration, we defined the first three, which refer to the evaluation of an existing article, as primary outcomes and the last one, the accuracy of a written headline, as a secondary outcome. This was due to the fact that we ex-ante believed to be more likely to find effects in a structured setting such as the evaluation tasks. In the following, we describe how these outcomes are measured.

Accuracy of the headline This variable is coded as 1 if the headline written by a journalist is factually accurate and 0 otherwise. Headline accuracy is assessed by research assistants who read the original scientific studies and then evaluate whether each headline aligns with the findings of the respective study. To account for potential subjectivity in classification, we rely on two independent human coders and assess inter-rater reliability using Cohen's kappa. The resulting agreement rate is 99.42 percent ($\kappa = 0.9885$), indicating a very high level of consistency and giving us confidence that the results are not driven by individual coders. As an additional robustness check, we prompted *ChatGPT* (o3 model) with the original studies and asked it to classify the headlines; this automated classification showed substantial overlap with human ratings. However, given that human assessments are typically regarded as the benchmark in evaluating classifications by large language models, we rely on the human-coded evaluations in our main analyses presented in Section 4.4.

Presence of errors This is a binary indicator equal to 1 if a participant's classification of an article's accuracy matches our pre-registered assessment, and 0 otherwise. Specifically, the indicator takes the value of 1 if the journalist agrees with our evaluation that the article either does or does not contain factual errors, and 0 if their assessment diverges from ours. As such, this variable is closely related to the "discernment" measures used in the misinformation literature (Guay et al., 2023), which capture whether individuals can correctly distinguish between true and false information. In our context, it reflects whether journalists are able to recognize when an article is accurate as well as when it contains errors.

Absolute number of errors This variable is an integer representing the absolute difference between the number of errors a journalist identified in a given article and the number of errors we pre-registered for that article. To reduce the influence of extreme outliers, we winsorized the variable at the 95th percentile.

Error types This is a binary indicator equal to 1 if the types of errors identified by a journalist match the error types we pre-registered for the respective article, and

0 otherwise.

4.3.1.4 Outcomes for Additional Analyses

In addition to our pre-registered outcomes we consider a set of additional outcomes. In particular, we (i) compute a Z-score index out of all pre-registered outcomes, (ii) re-code our data on the presence and absolute number of errors to get a measure of skepticism, (iii) classify our headline data in different linguistic dimensions and (iv) collect additional data on the articles written by participants after taking part in our study. This section describes how these additional variables are defined.

Joint effect on pre-registered outcomes To get an estimate of the overall effect of our treatment on all four pre-registered outcomes, we combine them to an standardized Z-score index. For this, we standardize each of the outcomes to have a mean of zero and a standard deviation of one. We multiply the absolute distance to the pre-registered number of mistakes with -1, as we want a higher value to be better for all of the values. We then compute the mean of these four standardized outcomes, giving equal weight to each one.

Skepticism: Article accuracy To capture how skeptical journalists were towards the articles they are exposed to, we considered whether they classify an article as containing a mistake or not (independent of whether this classification matches what we pre-registered). This is a dummy variable equal to 1 if a journalist classified a specific article as containing at least one mistake and 0 otherwise.

Skepticism: Number of mistakes Similar to the skepticism towards the article's accuracy we also considered the stated number of mistakes per article as a second measure of skepticism. This value is a positive integer. As this measure is potentially strongly influenced by outliers we winsorized it at the 95th percentile.

Linguistic dimensions of written headlines To further understand what kind of headlines participating journalists wrote, we enriched our data by classifying the headline text in different linguistic dimensions. We used OpenAI's GPT-4o model for the classification tasks. Each headline was sent to the model together with a concise system prompt that spelled out the scoring rubric for every variable. The exact prompt is denoted in Appendix C.7.2. The variables are: *sentiment* ($-1/0/1$), *emotionality*, *specificity*, *newsworthy*, *clickbait*, *density* (all continuous $[0, 1]$ scales), and an *extreme* language dummy. A detailed description on how every of these dimensions is defined is denoted in Table C11.

Subsequent reporting on science and scientific studies As outlined above, participants were invited to voluntarily disclose their names and the media outlets they write for at the end of the survey. Approximately one-quarter of participants (64 journalists) consented to do so. This information was stored separately from the main dataset and used to compile a database of articles published by each journalist in the two months following survey completion. We included up to 10 articles per journalist; if more than 10 were available, we selected the 10 published closest in time to their participation. If fewer than 10 articles were available, all were included. Only articles that were accessible through online search and freely available without a paywall were included. This procedure yielded a dataset of 268 real-world articles.

Research assistants then classified these articles along two dimensions: first, a binary variable indicating whether a scientific study was the main topic of the article (1 = yes, 0 = no); and second, a binary variable indicating whether a scientific study was cited at any point in the text (1 = yes, 0 = no). These classifications were then linked to the main dataset, resulting in two outcome variables for journalists who provided identifying information: (i) the total number of post-participation articles in which a scientific study was the main topic, and (ii) the total number of articles that cited a scientific study.

For the subset of articles that either focused on or cited a scientific study, we used *ChatGPT*'s deep research function (o3 model) to identify the original studies and assess whether any factual errors were present in their descriptions. Specifically, we prompted the model to classify 18–20 articles at a time and repeated the classification of each unique article three times (with memory disabled), resulting in a total of nine queries. Prompts used during this procedure are denoted in Appendix C.7. This approach, which involves multiple classification iterations, is similar to the procedure employed by Garg and Fetzer (2025) to evaluate the stability of *ChatGPT*'s outputs. We observed a high degree of agreement across iterations, with Cohen's κ values ranging from 0.63 to 0.83 (see Table C7). In cases where classifications differed across runs, we used the majority outcome as the final classification. This process generated a binary variable for each article in the subset, coded as 1 if one or more factual errors were identified in the article's representation of the study, and 0 otherwise.

4.3.2 Data

Our dataset comprises 260 responses from professional journalists in Germany, collected between August 13 and August 31, 2024. The median survey completion time was 22 minutes. In addition to the outcome measures described above, we collected data on participants' age, gender, political orientation, and income range, which are used as control variables in the analysis of our main outcomes. To assess balance across treatment conditions, we conducted pairwise t-tests or Chi-squared tests for these covariates between the treatment and control groups. Only one test indicated a statistically significant difference at the 10 percent level: a slightly higher proportion of participants in the treatment group reported earning *less than €1,500* compared to the control group ($p = 0.067$). As treatment was randomly assigned by computer, this minor imbalance is attributable to chance. We address it by including the full set of covariates as controls in our main regression models. The complete balance table is provided in Appendix C.1.1.

4.3.3 Statistical Analyses

We estimated the effects of our educational intervention using regression-based methods at the level of individual articles produced by journalists. Our primary approach compared outcomes between the treatment and control groups, controlling for relevant covariates such as demographic characteristics. To account for the repeated observations per journalist (since each journalist produced evaluations of two articles) we clustered standard errors at the journalist level to handle within-subject correlation.

Recognizing that not all participants fully engaged with the intervention (e.g., some may have skipped the video), we complemented our primary analysis with instrumental variable (IV) techniques. Here, assignment to the treatment group was used as an instrument to isolate the causal effect of reasonably long exposure to the educational video, thus addressing potential biases from partial compliance.

For follow-up analyses involving journalists' published articles after the intervention, we again used instrumental variable regression models and adjusted for clustering. We used IV due to the possibility that some control group members accessed the video post-survey. All technical details, including model specifications, variable definitions, and estimation procedures, are provided in Appendix C.4.

4.4 Results

In this section we present the results of our survey experiment. We start by illustrating the effect of our educational video on our pre-registered outcomes in section 4.4.1 and continue to explore its effects on the additional outcomes in section 4.4.2.

4.4.1 Main Results

Accuracy of written headlines Being exposed to the educational video significantly improves the most practically relevant outcome in our experiment: the quality of journalists' written work. Journalists in the treatment group are, on average, 28 percentage points more likely to produce an accurate headline compared to those in the control group ($p < 0.001$), as shown in Appendix Figure C1. The size of this effect corresponds to 0.6 standard deviations in the control group. Regression estimates are reported in Columns 1 and 2 of Table 4.1. The relatively large effect size is likely attributable to the design of our study. Specifically, it might be related to the selection of studies, which provided ample opportunity for the intervention to improve headline accuracy. A discussion of this aspect and its implications for external validity can be found in Section 4.5.2.

Presence of errors Participants in the treatment group were marginally more likely to match our pre-registered classification of an article's factual correctness, with a difference of 2 percentage points relative to the control group. However, this difference was not statistically significant ($p = 0.596$). Regression results are reported in Columns 3 and 4 of Table 4.1. Notably, participants in both groups performed no better than chance, with approximately 50 percent of classifications aligning with our benchmark in each group. This may indicate that the classification task was inherently difficult, potentially limiting participants' ability to detect factual inaccuracies as pre-registered reliably.

Error types Journalists in the treatment group were approximately 3 percentage points more likely to correctly identify the pre-registered type of error, although this difference was not statistically significant. Overall detection rates for pre-registered error types are relatively low: 46 percent in the control group and approximately 49 percent in the treatment group. Regression results for this outcome are reported in Columns 7 and 8 of Table 4.1.

When analyzing detection rates by error type, we observed a statistically significant improvement in the treatment group for errors related to statistical interpretation. Participants in the treatment group were 29 percentage points more likely to identify such errors when they were pre-registered as present ($p = 0.016$). Effect sizes by error type are visualized in Appendix C2.

Absolute number of errors The deviation from our pre-registered classification of the article-level errors was marginally greater in the treatment group than in the control group. On average, treated journalists deviated by 0.25 additional errors compared to control group participants ($p = 0.044$), as shown in Columns 5 and 6 of Table 4.1. Notably, this effect runs counter to expectations.

This finding suggests that the intervention does not enhance alignment with our pre-registered classifications. Instead, it could prompt journalists to adopt a more critical stance when evaluating others' reporting. For example, participants may have flagged statements as erroneous that our research assistants (who followed the standardized coding protocol) classified as accurate. One likely source of discrepancy lies in the subjective interpretation of concepts such as missing context: While our coders followed a detailed codebook, such structured guidance could not feasibly be provided to participating journalists due to time constraints.

Table 4.1: Main Results: OLS Estimates

	Written Headline		*Classification*		*Mistakes*		*Type*		*Index*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
TG	0.289***	0.292***	0.024	0.018	0.253**	0.231*	0.029	0.029	0.125**	0.128**
	(0.044)	(0.044)	(0.045)	(0.043)	(0.125)	(0.126)	(0.055)	(0.056)	(0.057)	(0.057)
Age		0.003		0.002		0.002		0.002		0.003
		(0.002)		(0.002)		(0.005)		(0.002)		(0.002)
Male		-0.025		0.039		-0.202		-0.037		0.025
		(0.045)		(0.045)		(0.132)		(0.059)		(0.059)
FE	no	yes	no	yes	no	yes	no	yes	no	yes
Constant	0.359***	0.120	0.500***	0.285	1.415***	1.870***	0.463***	0.444*	-0.060	-0.378*
	(0.031)	(0.163)	(0.032)	(0.167)	(0.082)	(0.337)	(0.039)	(0.196)	(0.040)	(0.219)
R^2	0.083	0.112	0.001	0.069	0.008	0.029	0.001	0.041	0.012	0.052
N	520	520	520	520	520	520	520	520	520	520

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the individual journalist in parentheses (260 clusters). Each column corresponds to a separate regression model. *Written Headline* refers to the accuracy of the written headlines. *Classification* refers to the accuracy of the classification of the provided news article. *Mistakes* refers to the absolute distance to the pre-registered number of mistakes in a provided news article. *Type* refers to the accuracy of the classification of the error type. *Index* is the Z-Score Index that combines all pre-registered outcomes. Fixed effects (FE) mean controls for the income level and political orientation of the participating journalists. *** p<0.01, ** p<0.05, * p<0.1.

Therefore, the apparent ineffectiveness of the treatment in improving this outcome (as well as the binary classification of error presence) may reflect limitations in instruction rather than a lack of intervention efficacy. Importantly, the fact that treated journalists performed significantly better when composing headlines suggests that the intervention supports a more critical and accurate approach to journalistic writing, even if it leads to divergence from our error classifications.

4.4.2 Additional Analyses

4.4.2.1 Joint effect on pre-registered outcomes

To assess overall performance across the pre-registered dimensions, we constructed a standardized index by calculating the Z-score of each outcome and aggregating all four pre-registered measures (see Section 4.3.1.4 for details on the computation). The resulting index was 0.12 standard deviations higher in the treatment group compared to the control group ($p = 0.029$). Regression results are reported in Columns 9 and 10 of Table 4.1. This positive effect was primarily driven by the substantial improvement in headline accuracy among treated participants.

4.4.2.2 Skepticism

To examine whether the unexpected results regarding the absolute deviation from our pre-registered error counts were driven by increased skepticism in the treatment group, we analyzed participants' error classifications without comparing them to our pre-registered benchmarks. This suggests that the intervention does indeed increase skepticism toward the articles. Treated journalists are 10 percentage points more likely to classify an article as (partially) factually incorrect ($p = 0.019$). In addition, they identified, on average, 0.45 more mistakes per article ($p = 0.001$). These differences are illustrated in Figure 4.2.

Taken together with the absence of consistent improvements in benchmark-based outcomes (Section 4.4.1), these findings indicate that the intervention primarily fostered generalized skepticism rather than improved discernment. In other words, journalists became more likely to flag potential errors, but not necessarily more accurate in distinguishing between correct and incorrect reporting from others.

4.4.2.3 Heterogeneity

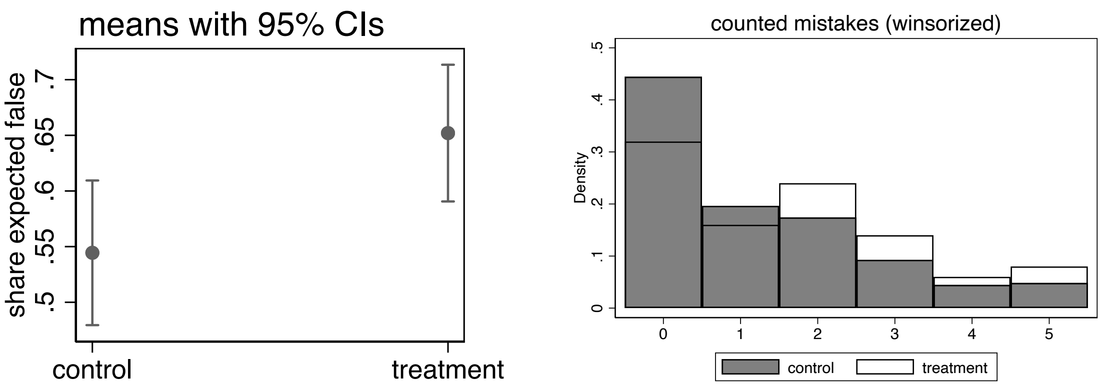
To explore potential heterogeneity in treatment effects across participant characteristics, we constructed binary indicators for each covariate and interacted them with the treatment variable in regressions on the pre-registered outcomes. To categorize participants by age, we defined a binary variable *Old*, equal to 1 for individuals aged 42 or older (the median age in our sample) and 0 otherwise. For income, *High Earner* was set to 1 for participants whose self-reported income falls within or above the median income category ("€2,500–€3,500"). For political orientation, *Left* equals 1 if a participant reports an intention to vote for the SPD, Die Linke, or Die Grünen, and 0 otherwise.

We found no evidence of systematic heterogeneity in our main results with respect to age or gender. Similarly, we did not observe significant heterogeneity by income or political orientation for most outcomes. However, for the accuracy of written headlines, we detected weakly statistically significant heterogeneity by income: the treatment effect was smaller for higher earners ($p = 0.073$), likely due to their higher baseline performance. Additionally, we find weak evidence of heterogeneity by political orientation for the absolute deviation from our classified number of mistakes ($p = 0.057$), suggesting that this result was primarily driven by left-leaning journalists. As the observed heterogeneities were only weakly statistically significant and not consistent across outcomes, we recommend interpreting them with caution and do not consider them a central finding of our study. Full regression results for these heterogeneity analyses are presented in Table C4 in Appendix C.1.4.

We also assessed whether clicking on detailed study materials influenced the efficacy of our treatment. Overall, only a small fraction of participants accessed the full study (15.4 percent in the untreated group and 17.5 percent in the treated group). This difference was not statistically significant in a Chi-squared test ($p = 0.550$). Therefore, treatment effects are unlikely to be linked to systematic heterogeneity in click-through behavior.

Figure 4.2: Results: Skepticism towards articles

(a) Means and 95%-confidence intervals **(b)** Distribution



Notes: Figure (a) illustrates the means and 95-percent confidence intervals of the share of articles classified as containing at least one mistake in each group. The estimates displayed here are obtained by running the baseline regression as described in equation C.1 (without controls) with robust standard errors clustered on the level of the journalist (260 clusters). Figure (b) depicts the distribution of the counted mistakes per article. As this measure is prone to be strongly influenced by outliers, it was winsorized at the 95th-percentile. Values for the control group are shaded in gray.

4.4.2.4 Linguistic dimensions of written headlines

To assess whether participants simply modified existing headlines, we computed cosine similarity between the original and rewritten versions. Cosine similarity measures the angle between two text vectors: a value of 0 indicates no lexical overlap, higher values reflect stronger similarity. Around 80 percent of rewritten headlines score 0. The mean similarity was 0.113 in the control group (SD = 0.175) and 0.104 in the treatment group (SD = 0.179). The difference was not statistically significant ($p = 0.59$). These results show that the treatment effects are not purely driven by superficial edits of the existing headlines. Appendix Figure C4 provides a histogram of the similarity distribution while Appendix Table C13 reports differences between our groups in cosine results for the headlines.

Beyond similarity, we also analyzed linguistic dimensions of rewritten headlines, namely sentiment, emotionality, specificity, newsworthiness, clickbait and density. Treated journalists wrote headlines that are on average more positive and less emotional than those in the control group. We did not find statistically significant differences between the headlines in the two groups in any other linguistic dimension we considered. Detailed regression results for every dimension are denoted in Table C12 in Appendix C.1.12.

4.4.2.5 Follow-Up: Analysis of real articles produced after participation

Approximately one quarter of our participants ($N = 64$) consented to sharing their names and employers for the purpose of analyzing their subsequent reporting. As outlined in Section 4.3.1.4, we used this information to compile a database of up to 10 of the most recent articles each journalist published within two months following their participation in our study. This procedure yielded a dataset of 268 real-world news articles.

We instructed research assistants to classify these articles along two dimensions: (i) whether the article primarily discusses a scientific study, and (ii) whether it cites a scientific study (e.g., to support or challenge a factual claim). The results revealed that scientific studies play a limited role in the reporting of most journalists in our sample. In 84 percent of cases, journalists did not publish any article in which science was the primary topic. Furthermore, 67 percent never cited a scientific study in these articles. This pattern is consistent with findings from prior content analyses on the limited presence of science in journalistic reporting (Suleski and Ibaraki, 2010; Summ and Volpers, 2016; Vestergaard and Nielsen, 2016).

To assess potential selection into the follow-up sample, we analyzed attrition in two stages: consent to name sharing and, conditional on consent, the likelihood of publishing science-related articles. Treated participants were more likely to consent to name sharing, but treatment assignment did not predict whether journalists write about scientific studies, either in the full sample or conditional on consent (Appendix Table C8). In addition, we examine whether performance and demographics differ systematically between respondents who agree to participate in the follow-up survey and those who decline. We do not detect any statistically significant differences (see Appendix Table C9). This indicates that while the follow-up sample is selective in levels, selection into published science articles is not differential by treatment. Accordingly, we interpret the follow-up results as suggestive evidence for a selected subpopulation.

To explore potential priming effects of our intervention, that is, whether watching the educational video increased the likelihood that journalists write about science or cite a scientific study, we estimated these outcomes using the IV approach described in Section C.4.3. On average, treated participants wrote 0.21 more articles with a scientific topic and cite a scientific study in 0.14 more articles than control group participants, although these differences were not statistically significant. The distributions of articles that discuss or cite scientific studies are shown in Figure C3 in Appendix C.2.

Given that the most plausible mechanism of our intervention was an improvement in the accuracy of journalistic reporting, we further analyze the subset of articles that either focus on a scientific study or cite one at any point (this is the second-stage subset from the attrition analysis). Using *ChatGPT*'s deep research function, we evaluated factual accuracy of these articles.

In this smaller subset, comprising 58 articles written by 21 journalists, we found that the vast majority are factually accurate, with only 14 percent containing at least one mistake. Applying the IV estimation approach described in Section C.4.3, we found that our intervention reduced the likelihood that a real-world article contains at least one factual error ($p = 0.020$). However, given the small size and likely selective nature of this sample, we recommend interpreting this exploratory finding with caution. To assess the robustness of the result, we repeated the IV estimation using each individual classification iteration produced by *ChatGPT*, instead of relying solely on the majority classification. While the direction of the effect remains consistent across iterations, the estimates were not statistically significant in every case. Detailed regression results are presented in Table C10 in Appendix C.1.10.

4.4.2.6 Robustness Checks

Correcting for Multiple Hypothesis Testing Given that our analysis evaluates the impact of the intervention on four pre-registered outcome variables, concerns regarding multiple hypothesis testing may arise. To account for this, we applied a Bonferroni correction to adjust our p-values accordingly. After the adjustment, only the effect on headline accuracy retains statistical significance at conventional thresholds. This underscores the robustness of this particular finding despite our relatively small sample size. This further supports our conclusion that improvements in headline accuracy constitute the most consistent and meaningful finding of our study. Table C5 in Appendix C.1.5 reports both the unadjusted and Bonferroni-corrected p-values for all pre-registered outcomes.

IV Analysis When analyzing our main outcomes using the IV estimation strategy as described in section C.4.2, all coefficients roughly double in size. In particular, treated journalists are 57 percentage points more likely to write an accurate headline ($p < 0.001$). Our Z-Score Index increased by 0.24 standard deviations for treated participants ($p = 0.030$). These findings indicate that a more intense treatment (i.e. spending more time watching our video) increased the effectiveness of the intervention. Detailed regression results of the IV estimations for all pre-registered outcomes and the Z-Score Index are denoted in Table C3 in Appendix C.1.

Headline Classifiers To evaluate the degree of subjectivity within in the classifications of headline accuracy, we compute Cohen's kappa comparing our main classification (i) with an independent human classifier and (ii) classifications generated by *ChatGPT*. This reveals a 99.42 percent agreement ($\kappa = 0.9885$) between the two human classifiers and a 64 percent agreement between the first human classifier and *ChatGPT* ($\kappa = 0.2816$). The lower agreement with *ChatGPT* is primarily driven by *ChatGPT* detecting a lower amount of mistakes in headlines than the humans.

As a robustness check, we replicated our analysis of the main result using the two alternative classifications. While the point estimates vary, the qualitative findings remained consistent and highly statistically significant regardless of the classifier used. Specifically, when relying on the classifications of the second human coder, assignment to the treatment group increased the likelihood of producing an accurate headline by 29 percentage points ($p < 0.001$), which corresponds to 0.61 standard deviations in the control group. Using *ChatGPT*'s classifications, the estimated effect was slightly smaller but still substantial, with treatment increasing the probability of writing an accurate headline by 17 percentage points ($p < 0.001$, the size of the difference corresponds to 0.33 standard deviations in the control group). Detailed regression results for these robustness checks are reported in Table C6 in Appendix C.1.6.

Continuous Controls We re-estimated our heterogeneity analyses without dichotomizing key moderators. Specifically, we modeled age as a continuous variable and interacted treatment assignment with age (TG × age), and we retained political orientation in its multi-category form (party choice) rather than collapsing it into broad left–right camps. Across all outcomes, we found no evidence of treatment effect heterogeneity with respect to age: the TG × age interaction was small and statistically insignificant for each error-type measure as well as for the composite z-score (see Appendix Table C14, Column 5). Similarly, allowing for multi-category political orientation yielded no systematic differences in treatment effects across political groups (Appendix Table C15).

4.5 Discussion

4.5.1 Divergent Effects on Pre-Registered Outcomes

Our study reveals mixed effects across the pre-registered outcomes, with a substantial and statistically significant improvement in one key dimension: headline accuracy. By contrast, the remaining three outcomes show no statistically significant effects; one even displaying a negative trend. Importantly, this loses significance once adjustments for multiple testing are applied.

These findings highlight both the robustness and practical relevance of the headline accuracy result. Of the four outcomes, we view headline writing as the most essential, given its centrality to journalistic practice. In contrast, evaluating others' work is less integral to journalists' day-to-day responsibilities. Moreover, headlines are known to strongly influence public understanding, given their disproportionate impact on reader perceptions (U. K. H. Ecker et al., 2014). At the same time, the overall explained variance in our headline regressions was modest, reflecting substantial heterogeneity in headline quality that was driven by factors beyond the scope of a brief intervention.

Further, the mixed results on the remaining outcomes warrant further reflection. The evaluation tasks such as detecting pre-registered errors in articles may have been inherently more challenging, as they require nuanced judgments, which likely increased measurement noise and attenuated treatment effects on these outcomes. While our video intervention appears effective in improving headline composition, enhancing journalists' ability to detect factual errors may require more extensive training or clearer evaluative frameworks. Further, more generally, while the intervention targets accuracy and overstatement, it does not explicitly train journalists to communicate scientific uncertainty, an important dimension for future work.

4.5.2 External Validity

Effect Size for Headline Accuracy The effect size we observed for headline accuracy was substantial, especially considering the brevity of our intervention. Previous work by Mehmood et al. (2021) also reported large effects, however their study involved a highly structured and intensive training program. In contrast, our findings showed that even a short, targeted intervention can produce meaningful improvements in journalistic performance, particularly in the accuracy of headlines.

A likely reason for the sizable treatment effect is the specific design of our study materials. We intentionally included scientific studies that are especially prone to media misinterpretation. This targeted selection makes the task of accurate reporting more difficult by design, and thus may have amplified the potential for our intervention to generate improvements. The high baseline error rate in the control group underscores this: 64 percent of their headlines were classified as factually inaccurate. These conditions created considerable scope for the training to have an effect, and the observed effect size should be interpreted in light of this context.

Our findings are also consistent with previous research documenting widespread inaccuracy in science reporting. For instance, Oxman et al. (2022) conduct a meta-analysis of health news articles and find that a large share are misleading. They often omit critical information such as harms, costs, or absolute effect sizes, or present exaggerated claims. While direct comparisons are difficult due to differences in coding criteria, their estimates suggest that between 30 and 80 percent of health news articles are misleading in at least one key dimension. The proportion of inaccurate headlines we document in the control group falls within this range, further supporting the external relevance of our setting.

Real-World Follow-Up on Published Articles A key limitation of our exploratory follow-up is the small and selective set of published articles available for analysis. Linking survey responses to real-world bylines required explicit participant consent, which only a subset of journalists provided. Moreover, our sample was not restricted to dedicated science reporters, limiting the number of relevant science articles. Constructing a broader sampling frame via bylines or scraping work emails was not feasible due to legal and ethical constraints. We therefore view this follow-up as an informative but necessarily limited first step. Importantly, despite the restricted sample, the results consistently suggest that the intervention reduced factual mistakes in published reporting, underscoring the potential value of further studies with more comprehensive newsroom data.

Selection due to Invitation E-Mail In our study, participants were informed in the recruitment email that the research focused on science journalism, which may have introduced some degree of selection bias. Journalists with a preexisting interest in science or a greater willingness to engage with scientific topics may have been more likely to participate. This has two potential implications. On the one hand, such journalists may have paid closer attention during the study, leading us to overestimate the effectiveness of the intervention relative to a broader population of journalists. On the other hand, their stronger baseline competencies in science reporting may have reduced the potential for improvement, potentially causing us to underestimate the true effect. Although the demographic characteristics of our sample are similar to those in a prior study that did not reference science in the recruitment materials (L. Berger, 2025), we cannot fully rule out the possibility that our recruitment approach influenced the composition of the participant pool.

Newsroom Organization A relevant scope condition of our headline results concerns newsroom workflows. In many news organizations, headlines are formally decided by editors rather than by reporters. In our experiment, we did not restrict participation to journalists who routinely finalize headlines, as doing so would have substantially limited statistical power. Moreover, in practice, headline formation is typically an argumentative and iterative process: reporters propose draft headlines or framing options that editors then revise. Under time pressure, editorial choices often build directly on the reporter's initial framing. Further, editors deciding on headlines frequently have less subject-matter expertise than the author of the article. This implies that inaccuracies at the headline stage may arise not only from reporters' misunderstandings, but also from editorial decisions made without full familiarity with the underlying study. From this perspective, our findings highlight editors as a particularly promising target group for future training.

Headline sentiment and audience incentives The intervention shifted headline sentiment in a more positive direction. This result stands in tension with evidence from digital news markets showing that negative headlines attract more clicks (L. Berger, 2025; Robertson et al., 2023). Importantly, our experiment is not designed to test headline selection under click-based incentives, but rather to isolate how improved science literacy affects journalists' interpretation and framing of scientific findings. Increased positivity should therefore be understood as a by-product of reduced exaggeration or alarmism: when journalists more accurately assess effect sizes, uncertainty, and scope, headlines are less likely to overstate risks or imply unwarranted conflict. Whether such higher-quality headlines are adopted in practice ultimately depends on editorial priorities and incentive structures.

Cultural Context Our research is conducted within the context of German journalism, which possesses distinct structural and cultural characteristics. For example, Germany consistently ranks among the top countries on the World Press Freedom Index (Reporters Without Borders, 2024) and benefits from strong public broadcasting institutions, fostering a relatively favorable environment for journalistic rigor. However, our findings indicate that even in such conditions, journalists may perform poorly when reporting on scientific studies. This raises important questions about the external validity of our findings, particularly in media environments with weaker press freedoms, more constrained resources, or differing audience expectations. If journalistic shortcomings in science reporting persist despite Germany's institutional advantages, they may be even more pronounced in settings where structural challenges further limit journalists' capacity for accurate and nuanced reporting.

4.6 Conclusion

This study shows that a brief, low-cost intervention can meaningfully improve journalists' science literacy. By focusing on core principles of statistical interpretation and science communication, our short video markedly improved the accuracy of headlines — a central and highly visible component of journalistic work. These findings underscore the potential of scalable, evidence-based training tools to improve the quality of science reporting and promote more accurate, trustworthy media coverage of scientific topics.

Our results also demonstrate broader challenges journalists face in critically engaging with scientific studies. While Germany provides a relatively favorable media environment, characterized by strong public broadcasting institutions and high press freedom, journalists in our sample frequently misinterpreted scientific findings and struggled with assessing errors of others. This suggests that deficits in science literacy persist even in structurally advantageous settings, reinforcing the need for continued professional development in this area.

Journalists play a crucial role as intermediaries between science and the public. They act as multipliers, shaping public understanding by embedding scientific findings into accessible and engaging narratives. This role can be both beneficial and problematic: While journalists can increase public science literacy by drawing attention to important research, they may also contribute to public misunderstanding by disseminating inaccurate or exaggerated claims. Improving science literacy among journalists thus offers a cost-effective strategy for maximizing the benefits of science communication while minimizing potential harms.

Our findings have direct implications for professional education and training in journalism. By demonstrating that a brief, cost-effective intervention can significantly improve the accuracy of science reporting, we present a scalable model for professional development in journalism. News organizations and journalism schools should consider integrating similar training modules into professional development programs and journalism curricula. Furthermore, funding agencies and media regulators could incentivize the adoption of such interventions through grants or accreditation programs, acknowledging their potential to enhance the societal value of journalism. By equipping journalists with better tools to critically engage with scientific findings, these measures can mitigate misinformation, improve reporting accuracy, and ultimately strengthen public trust in both science and the media.

Data Statement

The experiment was approved by the Ethics Committee of the University of Cologne (reference 240007LB) and pre-registered in the AEA RCT Registry under registry number AEARCTR-0014037. Financial support from the Joachim Herz Stiftung is gratefully acknowledged. A replication package is available on OSF, including the anonymized dataset used for all analyses, the analysis code, and the ChatGPT rating outputs: `https://doi.org/10.17605/OSF.IO/VQSZT`. To protect participant anonymity, we do not share real names, employers, or the original texts of the real-world articles used in the follow-up analysis and provide anonymized article identifiers and all corresponding classifications instead.

Bibliography for Chapter 4

Armona, L., M. Gentzkow, E. Kamenica, and J. M. Shapiro (2024). “What is Newsworthy? Theory and Evidence”. In: *Working Paper*. Forthcoming.

Balbuzanov, I., J. Gars, M. Stalinski, and E. Tjernström (2025). “Incentivizing Engagement: Experimental Evidence on Journalist Performance Pay”. In: Working paper.

Berger, L. (2025). “How digital media markets amplify news sentiment”. In: *Working Paper*.

Berger, L. M., A. Kerkhof, F. Mindl, and J. Münster (2025). “Debunking “fake news” on social media: Immediate and short-term effects of fact-checking and media literacy interventions”. In: *Journal of Public Economics* 245, p. 105345.

Cacciatore, M. A. (2021). “Misinformation and public opinion of science and health: Approaches, findings, and future directions”. In: *Proceedings of the National Academy of Sciences* 118.15, e1912437117.

Chahrour, R., K. Nimark, and S. Pitschner (2021). “Sectoral Media Focus and Aggregate Fluctuations”. In: *American Economic Review* 111.12, pp. 3872–3922.

Dahlstrom, M. F. (2021). “The narrative truth about scientific misinformation”. In: *Proceedings of the National Academy of Sciences* 118.15, e1914085117.

Dempster, E., R. Sutherland, and S. Keogh (2022). “Scientific Research in News Media: A Case Study of Misrepresentation, Sensationalism and Harmful Recommendations”. In: *Journal of Science Communication* 21.1, A06.

Durante, R. and E. Zhuravskaya (2018). “Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict.” In: *Journal of Political Economy* 126.3, pp. 1085–133.

Ecker, U. K. H., S. Lewandowsky, E. P. Chang, and R. Pillai (2014). “The effects of subtle misinformation in news headlines”. In: *Journal of Experimental Psychology: Applied* 20.4, pp. 323–335.

Eisensee, T. and D. Strömberg (2007). “News Droughts, News Floods, and U.S. Disaster Relief”. In: *The Quarterly Journal of Economics* 122.2, pp. 693–728.

Garg, P. and T. Fetzer (2025). “Political expression of academics on Twitter”. In: *Nature Human Behaviour*.

Graves, L., B. Nyhan, and J. Reifler (2016). “Understanding innovations in journalistic practice: A field experiment examining motivations for fact-checking”. In: *Journal of Communication* 66.1, pp. 102–138.

Guay, B., A. J. Berinsky, G. Pennycook, and D. Rand (2023). “How to think about whether misinformation interventions work”. In: *Nature Human Behaviour* 7.8, pp. 1231–1233.

Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). “A digital media literacy intervention increases discernment between mainstream and false news in the United States and India”. In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15536–15545.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). “Curtailing False News, Amplifying Truth”. In: *Working Paper*.

Hanitzsch, T., N. Steindl, and C. Lauerer (2016). “Country Report: Journalists in Germany”. In: *Worlds of Journalism Study*.

Kirkegaard, E. O. W., J. Pallesen, E. Elgaard, and N. Carl (2021). “The Left-liberal Skew of Western Media”. In: *Journal of Psychological Research* 3.

Markowitz, D. M. (2024). “From complexity to clarity: How AI enhances perceptions of scientists and the public’s understanding of science”. In: *PNAS Nexus* 3.9, p. 387.

Mehmood, S., S. Naseer, and D. L. Chen (2021). “Training policymakers in econometrics”. In: *Working Paper*.

Oxman, M., L. Larun, G. P. Gaxiola, D. Alsaid, A. Qasim, C. J. Rose, K. Bischoff, and A. D. Oxman (2022). “Quality of information in news media reports about the effects of health interventions: Systematic review and meta-analyses”. In: *F1000Research* 10, p. 433.

Rauscher, F. H., G. L. Shaw, and K. N. Ky (1993). “Music and spatial task performance”. In: *Nature* 365.6447, p. 611.

Reporters Without Borders (2024). *World Press Freedom Index 2024*.

Robertson, C. E., N. Pröllochs, K. Schwarzenegger, P. Pärnamets, J. J. Van Bavel, and S. Feuerriegel (2023). “Negativity drives online news consumption”. In: *Nature Human Behaviour* 7.5, pp. 812–822.

Roozenbeek, J., S. van der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky (2022). “Psychological inoculation improves resilience against misinformation on social media”. In: *Science Advances* 8.34, eabo6254.

Selvaraj, S., D. S. Borkar, and V. Prasad (2014). “Media coverage of medical journals: Do the best articles make the news?” In: *PLOS ONE* 9.1, e85355.

Serra-Garcia, M. (2025). *The Attention–Information Tradeoff*. CESifo Working Paper 11885. Available at SSRN. CESifo.

Suleski, J. and M. Ibaraki (2010). “Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media”. In: *Public Understanding of Science* 19.1, pp. 115–125.

Summ, A. and A.-M. Volpers (2016). “What’s science? Where’s science? Science journalism in German print media”. In: *Public Understanding of Science* 25.7, pp. 775–790.

Sumner, P., S. Vivian-Griffiths, J. Boivin, A. Williams, L. Bott, R. Adams, C. A. Venetis, L. Whelan, B. Hughes, and C. D. Chambers (2016). “Exaggerations and Caveats in Press Releases and Health-Related Science News”. In: *PLoS ONE* 11.12, e0168217.

Vestergaard, G. L. and K. H. Nielsen (2016). “Science news in a closed and an open media market: A comparative content analysis of print and online science news in Denmark and the United Kingdom”. In: *European Journal of Communication* 31.6, pp. 661–677.

Willnat, L., D. H. Weaver, and G. C. Wilhoit (2022). “The American Journalist Under Attack: Key Findings 2022”. In: *Report S.I. Newhouse School of Public Communications*.

Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. 1st. London: Atlantic Books.

Appendix A

Appendix to Chapter 2

A.1 Supplementary Materials Descriptives

A.1.1 Description of Sentiment Classifiers

The numerous classifiers available can be sorted into machine learning approaches and lexicon-based techniques. While the most recent machine learning classifiers oftentimes produce a higher accuracy (bench-marked with human-coded data), lexicon-based classifiers are much less of a black-box in their classification decisions.

A.1.1.1 Dictonary-based approaches

For the dictonary-based classifications I remove stopwords from the headlines using stopword dictionaries and stem the headlines in a first step[1]. Then, the number of positive and negative words in each headline is counted. The *sentiment score* is then the number of positive words minus the number of negative words, which is a standard way to express the overall tonality of textual data with dictionary approaches. To enable a comparison with the human coded classifications I then define a headline to be positive if the sentiment score is larger than zero, negative when it is smaller than zero and neutral when it is equivalent to zero.

SentimentWortschatz (SentiWS)

The SentimentWortschatz (SentiWS) is a publicly available German language resource for sentiment analysis. With around 1,650 positive and 1,800 negative basic forms (which results in around 16,000 positive and 18,000 negative word forms

[1]Stopwords are words that usually do not contain any meaning such as “the” or “a”. They are removed to decrease computation time. For the English headlines I use the stopword-dictonary from the R tidytext package. For the German headlines I use a stopwords list from the countwordsfree blog.

including the various inflected forms) it is to the best of my knowledge the largest German-language sentiment dictionary.

LoughranMcDonald-dictonary (LM)
The LoughranMcDonald-dictonary (LM) is a widely used English-language finance-specific dictionary. As I consider news articles on economic issues this dictonary might be a better fit than general sentiment dictionaries.

Valence Aware Dictionary and sEntiment Reasoner (VADER)
A common mistake in sentiment classifications by dictionary approaches happens when valence shifters are not taken into account. For example the sentence “I am not happy” would be classified as positive, as it contains the word “happy”. This problem is addressed by the Valence Aware Dictionary and sEntiment Reasoner (VADER). This algorithm is a lexicon and rule-based sentiment analysis approach that in addition to word counts respects negations and other commonly used language rules. The likelihood of the above described mistakes is here therefore much lower, which could increase the accuracy of the classifications. VADER was however mainly created to classify social media content, which in turn might lower accuracy. The classifications of VADER are distributed on a scale from -2 (very negative) to +2 (very positive). Again, for the comparison with the human coded data I consider a headline to be positive if the VADER score is above zero, negative when it is smaller than zero and neutral when it is equivalent to zero.

A.1.1.2 Machine learning approaches

Financial-RoBERTa (pre-trained)
The machine learning technique I use first is called Robustly Optimized Bidirectional Encoder Representations from Transformers (RoBERTa). It is an optimized BERT pretraining approach which is currently often considered the state-of-the-art for text classifications (A. H. Shapiro et al., 2022). BERT itself is a self-supervised machine learning technique introduced by Google in 2018 for Transformer-based Natural Language Processing models. The model that I use is called Financial-RoBERTa and was pre-trained to analyze sentiment of financial texts. The training data included financial statements, earnings announcements, earnings call transcripts, corporate social responsibility reports, and news articles on environmental, social, governance and finance topics (Soleimanian, 2022). Financial-RoBERTa sorts the headlines into the three classes positive, negative or neutral which can be directly compared with the human-coded data.

Financial-RoBERTa (fine-tuned)
For further fine-tuning of Financial-RoBERTa I split the human-coded dataset (N=2500) into a training-, evaluation- and test-dataset[2]. The training data is used to train an additional layer of the Financial-RoBERTa model and the evaluation data is used to fine-tune the hyper-parameters of the training process. I use grid-search to fine tune the hyper-parameters and end up with a learning rate of 0.00002, 4 training epochs, a batch size per device during training of 16, a total number of 230 steps and 50 warm-up steps[3]. Again, this model sorts the headlines into the three classes which can be directly compared with the human-coded data.

GPT-3.5-Turbo
The second technique relies on *OpenAI*'s Generative Pre-trained Transformer (GPT) family of large language models (LLMs). GPT models are trained with a next-token–prediction objective on a broad mixture of web documents and books and are therefore able to perform many downstream language tasks without additional parameter updates. I use the public "gpt-3.5-turbo-0125" checkpoint, which at inference time is steered only through a natural-language system prompt that instructs the model to output a strict JSON object containing the sentiment label. In this zero-shot setting the model leverages its emergent in-context reasoning abilities to assign each headline to the classes *positive*, *neutral*, or *negative*.

GPT-3.5-Turbo (fine-tuned)
To further align GPT with the specific tone of economic news in my dataset I fine-tune the base model on the human-coded headline corpus ($N = 2500$).[4] Fine-tuning is performed via the *OpenAI* fine-tuning API, which adds a lightweight adaptation layer to the frozen base model. The job is run for the default three epochs with the adaptive learning-rate schedule chosen by *OpenAI*. Validation loss is monitored after every epoch and early stopping is triggered if the metric deteriorates. The resulting model outputs the integer-coded sentiments directly, so that the predictions can again be compared one-to-one with the human-coded labels.

[2]I follow the convention of using 70% of the data as training data, 15% as evaluation data and the remaining 15% as test-data. For the dataset at hand this results in 1750 training observations, 375 evaluation observations and 375 test observations.

[3]The seed was set to "1234".

[4]Following common practice, the corpus is divided into 70% training data, 15% evaluation data, and 15% test data, resulting in 1.750 training observations, 375 evaluation observations and 375 test observations.

A.1.1.3 Evaluation of the classifications

I use the test-dataset to evaluate and compare all of the classification of the different algorithms. Table A1 provides an overview of the accuracy and macro F1 scores of each model. In addition to the *sentiment* classifications it is interesting for my research question to know whether an algorithm is able to correctly identify emotional language independently of it being positive or negative. I therefore define an *emotonality* dummy which is equal to one if a headline was classified as positive or negative and zero otherwise and compute accuracy and F1 scores for this outcome as well.

Table A1: Evaluation of Sentiment Classification Algorithms

	Sentiment		*Emotionality*	
algorithm	**accuracy**	**macro F1**	**accuracy**	**F1**
SentiWS	0.4773	0.3729	0.5413	0.3174
LM	0.6106	0.5257	0.6373	0.5436
VADER	0.5920	0.5554	0.6800	0.6428
pre-trained roBERTa	0.6853	0.6638	0.7200	0.7301
fine-tuned roBERTa	0.7253	0.7057	0.7413	0.7581
GPT 3.5	0.7573	0.7567	0.7627	0.7601
fine-tuned GPT 3.5	0.7973	0.7867	0.8213	0.8337

Notes: Table A1 presents the accuracy and the (macro) F1 scores for sentiment and emotionality of the different sentiment classifiers on the test-dataset. Accuracy is the share of correct classifications, while the F1-score is calculated considering both the precision and recall of classifications. The macro F1 averages the F1 score of the different classes (positive, neutral, negative).

A.1.2 Comparisons with Controls

In addition to the baseline specification as described in equation 2.1, I run my descriptive analysis controlling subsequently for different potential drivers of the difference in tonality of online and offline headlines. The regressions that I run are described in Equations A.1 to A.3.

$$tonality_i = \beta_0 + \beta_1 online_i + \beta_2 content\_tonality_i + \epsilon_i \tag{A.1}$$

$$tonality_{ik} = \beta_0 + \beta_1 online_i + \beta_2 topic_k + \epsilon_{ik} \tag{A.2}$$

$$tonality_{it} = \beta_0 + \beta_1 online_i + \beta_2 time_t + \epsilon_{it} \tag{A.3}$$

$content\_tonality_i$ is a measure of sentiment or emotionality of the content of the article (not the headline). It is obtained by classifing the content with the same fine-tuned sentiment classifier as the headlines. For coherence this measure is equivalent to the contents' emotionality when I analyze emotionality of the headline as an outcome and equivalent to the contents' sentiment when the sentiment of the headline is regarded.

$time_t$ is a vector with dummies for each day in the data.

$topic_k$ is a categorical variable that contains the most prominent topic out of 21 possibilities. The considered topic categories are: International, Defense, Government, Civil Rights, Environment, Transportation, Law and Crime, Energy, Health, Domestic Commerce, Immigration, Labor, Macroeconomy, Agriculture, Social Welfare, Technology, Education, Housing, Foreign Trade, Culture, Public Lands. These categories are assigned to the headlines using the ParlBERT-Topic-German model (Klamm et al., 2022).

Equation A.4 describes a specification where I use all described covariates jointly and additionally control for article length[5] and agency content[6].

$$\begin{aligned} tonality_{ikt} = \beta_0 &+ \beta_1 online_i + \beta_2 content\_tonality_i \\ &+ \beta_3 topic_k + \beta_4 time_t + \beta_5 length_i + \beta_6 agency_i + \epsilon_{ikt} \end{aligned} \tag{A.4}$$

$length_i$ is a numerical variable with the number of words in the article that belongs

[5]This control is added as longer reads might be edited differently than quickly produced stories.

[6]This control is added because the process of choosing a headline for a pre-written story by a news agency might be different from the one for an article that was written by a journalist of the publishing outlet.

to the headline.

agency$_i$ is a dummy variable equal to 1 if a news agencies name is included in the author line and 0 otherwise.

In Equation A.4 the β_1 thus corresponds to the difference in tonality between online and offline headlines independent of the tonality of the article, time of publication, topic, agency-content and length. If there is still a difference detectable an explanation could be that journalists frame their stories differently for the online sphere, i.e. emphasize the emotional aspects of a story more often in the headline.

A.1.3 Matching of Identical Articles

The objective is to identify near identical articles that appear both online and offline within the same outlet in order to compare sentiment and emotionality holding content constant. Each article has a unique identifier `ID` and metadata that include outlet and publication channel.

Text Cleaning

All text is normalised as follows before any similarity computation. First, it is converted to lower case. Then, punctuation is removed and multiple whitespaces are collapsed to a single space. Finally, missing strings are replaced by the empty string so that row order and identifiers remain aligned.

Candidate generation with five word shingles and locality sensitive hashing

To avoid a quadratic all against all comparison, which would have been extremely computationally heavy, I first generate a compact candidate set for each article using five word shingles and MinHash with banding in the following procedure:

1. Tokenisation into overlapping five word sequences. For article i with token sequence $(w_1, \ldots, w_T)$, create the multiset of shingles $\{(w_t, \ldots, w_{t+4})\}_{t=1}^{T-4}$.

2. MinHash signatures. Use $n_{\text{perm}} = 256$ independent permutations to map each shingle multiset to a signature vector $m_i \in \mathbb{N}^{256}$ that preserves the Jaccard similarity in expectation.

3. Locality sensitive hashing with banding. Partition each signature into $b = 64$ bands with $r = 4$ rows per band. Two articles become candidates if they share at least one identical band hash. This configuration concentrates recall on Jaccard similarity above roughly 0.8.

4. Candidate set extraction. Collect unordered pairs (i, j) that collide in at least one band, drop self pairs, and drop exact duplicates of the same pair. This reduces the average fan out to a few dozen candidates per article.

Semantic re–scoring with sentence level embeddings

The candidates from the hashing stage are then evaluated with a more expressive and precise similarity measure that captures paraphrases and reordering of sentences. Every article is embedded into a dense numerical vector using a pretrained Sentence Transformer model. For this, I use the model *all MiniLM L6 v2* which yields a three hundred eighty four dimensional vector for each document. All vectors are scaled to unit length so that cosine similarity equals the simple dot product of two vectors. Cosine similarity is then computed for every candidate pair from the hashing stage. This two stage design keeps computation linear in the number of documents while providing a high quality semantic score.

After this, the pairs are sorted. Only cross channel pairs within the same outlet are retained, meaning one member of the pair is published online and the other member is published offline. Self pairs are removed. If one article has several candidates on the opposite channel the default rule keeps the partner with the highest cosine similarity.

Thresholding and creation of matched subsets

Cosine similarity thresholds are used to translate scores into final matches. Five cutoffs are defined at 0.97, 0.95, 0.92, 0.90, and 0.85. For each cutoff a dataset is created that contains all pairs with a score at or above the cutoff. Each record stores the two article identifiers and their cosine similarity. These five datasets allow analyses that trade off match precision against match count in a transparent way.

A.1.4 Descriptive Data Summary Statistics

Table A2: Summary statistics of dependent variables, entire dataset

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
content emotionality	0.573	0.494	0	1	339,865
content sentiment	-0.289	0.699	-1	1	339,865
article length	0.753	0.688	0.001	71.382	339,865
agency content	0.381	0.485	0	1	339,865
Outlet					
Rheinische Post	0.099	0.299	0	1	339,865
BILD	0.023	0.149	0	1	339,865
Der Spiegel	0.275	0.446	0	1	339,865
Die Welt	0.264	0.440	0	1	339,865
Die Zeit	0.338	0.473	0	1	339,865
Topic					
Agriculture	0.022	0.148	0	1	339,865
Civil	0.084	0.278	0	1	339,865
Culture	0.005	0.076	0	1	339,865
Defense	0.015	0.122	0	1	339,865
Domestic	0.178	0.382	0	1	339,865
Education	0.006	0.081	0	1	339,865
Energy	0.022	0.148	0	1	339,865
Environment	0.030	0.172	0	1	339,865
Foreign	0.010	0.101	0	1	339,865
Government	0.223	0.416	0	1	339,865
Health	0.026	0.159	0	1	339,865
Housing	0.015	0.124	0	1	339,865
Immigration	0.008	0.091	0	1	339,865
International	0.115	0.319	0	1	339,865
Labor	0.024	0.155	0	1	339,865
Law	0.033	0.180	0	1	339,865
Macroeconomic	0.069	0.253	0	1	339,865
Public	0.002	0.048	0	1	339,865
Social Welfare	0.019	0.139	0	1	339,865
Technology	0.037	0.188	0	1	339,865
Transportation	0.046	0.209	0	1	339,865
Time					
2022	0.068	0.252	0	1	339,865
2021	0.127	0.333	0	1	339,865
2020	0.106	0.308	0	1	339,865
2019	0.085	0.279	0	1	339,865
2018	0.070	0.255	0	1	339,865
2017	0.057	.231	0	1	339,865
2016	0.046	0.210	0	1	339,865
2015	0.054	0.227	0	1	339,865
2014	0.053	0.225	0	1	339,865
2013	0.059	0.237	0	1	339,865
2012	0.053	0.224	0	1	339,865
2011	0.043	0.204	0	1	339,865
2010	0.044	0.206	0	1	339,865
2009	0.046	0.209	0	1	339,865
2008	0.018	0.134	0	1	339,865
2007	0.016	0.128	0	1	339,865
2006	0.013	0.115	0	1	339,865
2005	0.012	0.109	0	1	339,865
2004	0.011	0.106	0	1	339,865
2003	0.009	0.097	0	1	339,865

A.1.5 Descriptive Results

Table A3: OLS Estimates – Emotionality of Headlines

	(1)	(2)	(3)	(4)	(5)
online	0.1861***	0.1231***	0.1542***	0.1582***	0.1043***
	(0.0387)	(0.0230)	(0.0361)	(0.0295)	(0.0201)
content emotionality		0.3199***			0.2975***
		(0.0141)			(0.0112)
article length					0.0101
					(0.0087)
agency content					0.0205*
					(0.0122)
topic FE	no	no	yes	no	yes
time FE	no	no	no	yes	yes
Constant	0.4401***	0.2956***	0.4249***	0.000	-0.0270
	(0.0396)	(0.0098)	(0.0606)	(0.0191)	(0.0146)
R^2	0.0330	0.1306	0.0677	0.0171	0.1311
Observations	339,865	339,865	339,865	339,865	339,865

Notes: Table A3 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The offline headlines are always the reference group. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. *** p<0.01, ** p<0.05, * p<0.1

Table A4: OLS Estimates – Sentiment of Headlines

	(1)	(2)	(3)	(4)	(5)
online	-0.1021***	-0.0690***	-0.1041***	-0.0931***	-0.0759***
	(0.0349)	(0.0143)	(0.0273)	(0.0255)	(0.0120)
content sentiment		0.5316***			0.5219***
		(0.0238)			(0.0224)
article length					-0.0222
					(0.0217)
agency content					0.0062
					(0.0191)
topic FE	no	no	yes	no	yes
time FE	no	no	no	yes	yes
Constant	-0.1172***	0.0161	-0.1021	0.000	0.0196
	(0.0387)	(0.0303)	(0.0527)	(0.0168)	(0.0163)
R^2	0.0047	0.2688	0.0219	0.0028	0.2687
Observations	339,865	339,865	339,865	339,865	339,865

Notes: Table A4 reports OLS estimates with bootstrapped standard errors in parentheses. Bootstrapping is conducted with 50 replications based on 5 clusters at the level of the news outlet. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The offline headlines are always the reference group. Control variables are a content emotionality dummy, article length in 1000 words, a agency content dummy, outlet fixed effects as dummies for the different outlets, topics and points in time. *** p<0.01, ** p<0.05, * p<0.1

A.1.6 Robustness Checks

A.1.6.1 Time Trends: Overall picture

Examining how headline sentiment and emotionality evolve over time is informative, as it helps determine whether the observed differences are persistent or limited to specific periods. Moreover, understanding whether these differences have increased, decreased, or remained relatively stable in recent years allows for a more meaningful interpretation of the results.

Figure A1: Emotionality and Sentiment of Headlines over Time

(a) Emotionality **(b)** Sentiment

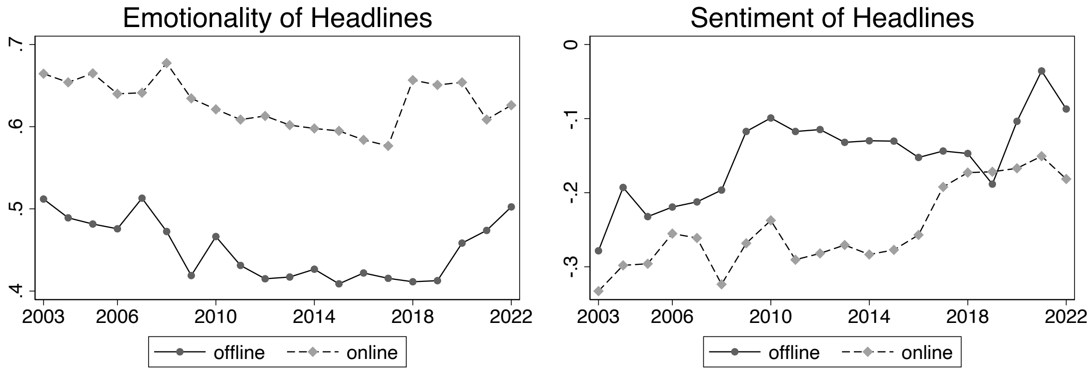
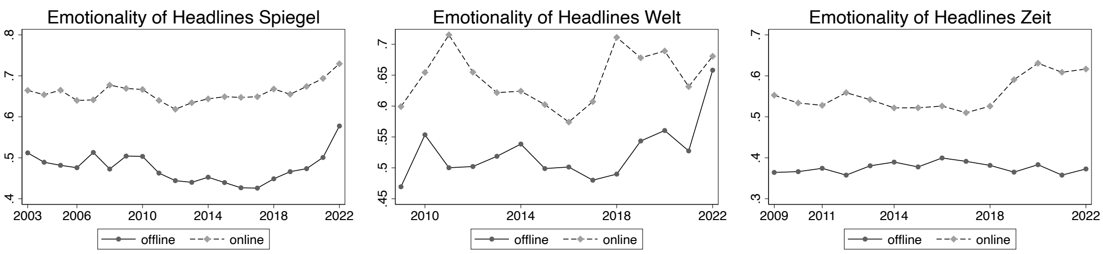
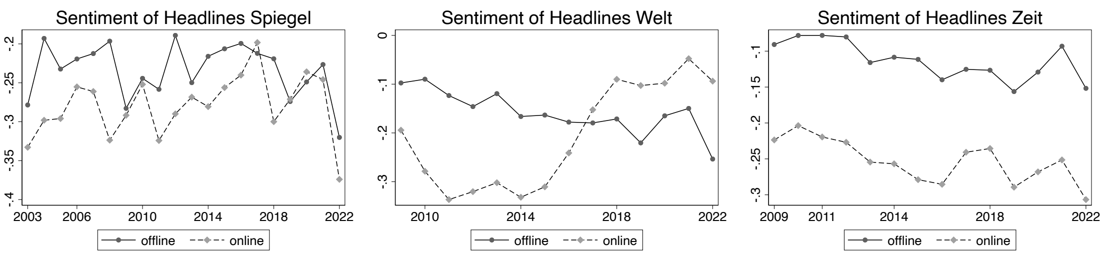


Figure A2: Emotionality and Sentiment of Headlines over Time

(a) Der Spiegel **(b)** Die Welt **(c)** Die Zeit



(d) Der Spiegel **(e)** Die Welt **(f)** Die Zeit



Notes: Figures A2(a) - A2(c) illustrate the development of emotionality over time. Figures A2(d) - A2(f) illustrate the development of sentiment over time. Values for the offline headlines are illustrated by a full line. The online values are depicted with a dotted line.

Figure A1 shows the evolution of emotionality and sentiment over time for the entire dataset. Figure A2 presents the same patterns for the three outlets with the longest time coverage. Overall, the difference in emotionality appears to be quite stable. It persists across all years and outlets, and no clear trend of increasing or decreasing emotionality is evident.

Online news are on average almost always more negative than offline news. However, the difference in sentiment appears to be more outlet-specific, as illustrated in Figures A2d–f. For *Der Spiegel*, the gap between online and offline sentiment seems to have narrowed over time. For *Die Zeit*, the difference remains relatively stable, with online headlines being on average more negative than their offline counterparts in nearly every year. In contrast, *Die Welt* followed a similar pattern until around 2015, but its online headlines have become considerably more positive since 2016 and are now, on average, even more positive than offline headlines. A more detailed discussion and potential explanation for this shift can be found in Appendix A.1.6.2.

Overall, the analysis of time trends shows that the difference in emotionality remains fairly stable over time. Although online headlines are on average more negative than their offline counterparts at most points in time, this difference appears to be more outlet-specific.

A.1.6.2 Time trends: Shift in Sentiment for *Die Welt*

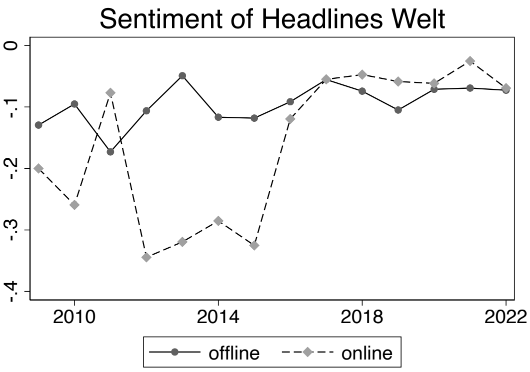
The analysis of time trends reveals an arguably surprising pattern: while online headlines of *Welt Online* were initially more negative than their offline counterparts, this relationship reversed in 2016, and online headlines have since become more positive. This development is not observed for the other outlets, raising the question of what drives this change. A closer examination indicates that the shift is entirely explained by an increased share of agency content. Figure A3 illustrates the evolution of sentiment separately for agency and non-agency material.

Examining the absolute number of agency articles published per month shows that, coinciding with the shift in sentiment, the volume of agency content increased sharply. Between October 2015 and October 2016, the average number of agency articles on economic topics per month was 42. In the same period one year later, the average rose to 215. Table A5 provides an overview of the number of online agency articles published by *Die Welt* in the main dataset between 2015 and 2017.

To understand the cause of this substantial increase in agency content, I examined archived versions of *welt.de* from this period using the *Internet Archive's Wayback Machine*. Interestingly, a notable change occurred in the *News Ticker* section on September 13, 2016. Before this date, the ticker served as an overview page that

Figure A3: Sentiment of Welt Headlines by Agency Content

(a) Agency Content



(b) No Agency Content

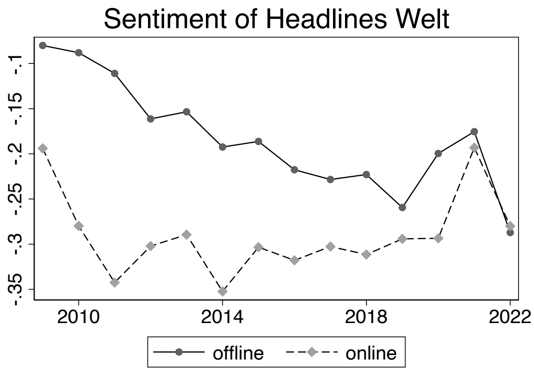


Table A5: Count of Online Agency Articles by *Die Welt* in Main Dataset

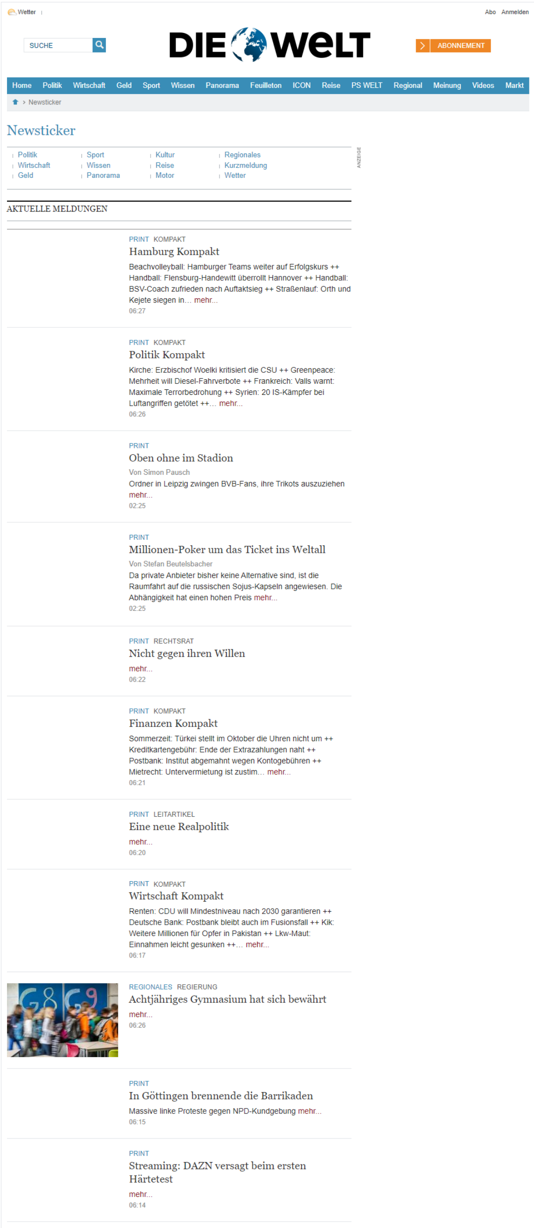
2015		**2016**		**2017**	
month	*N*	*month*	*N*	*month*	*N*
January	74	January	38	January	209
February	95	February	52	February	220
March	79	March	47	March	268
April	75	April	36	April	202
May	74	May	31	May	217
June	92	June	31	June	224
July	70	July	26	July	227
August	59	August	25	August	222
September	41	September	53	September	207
October	75	October	90	October	187
November	30	November	221	November	203
December	48	December	179	December	140

listed many articles based on agency content, alongside other types of articles such as direct copies from the print edition. It appeared to be used rather infrequently and still displayed an outdated website design, whereas the main landing page had already adopted a newer layout. From September 13 onwards, the ticker page also switched to the updated design and featured both a section summarizing the most important news and a continuously updated ticker that included a large volume of agency content.

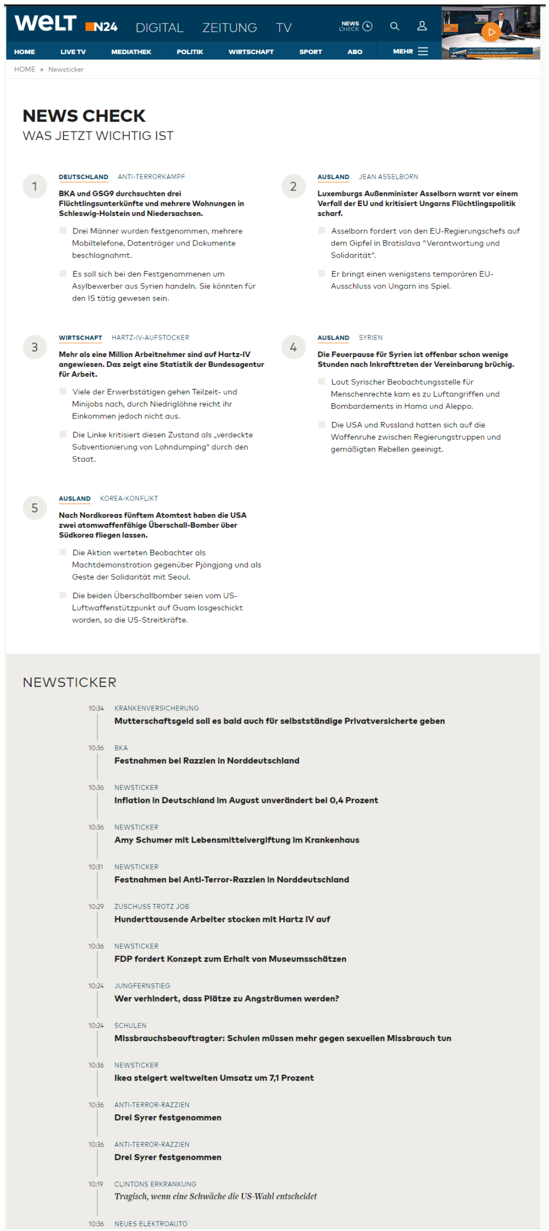
It is therefore plausible that the introduction of this new ticker format led to the automated publication of a much higher number of agency articles, thereby increasing the share of agency content in the online sample to an extent that it affected the average sentiment difference between online and offline headlines. Figure A4 displays screenshots of the ticker page before and after the redesign. The earlier version is available via the Wayback Machine here, and the newer version here.

Figure A4: News Ticker at Welt Online

(a) Before 13.09.2016



(b) After 13.09.2016



A.1.6.3 Matched Subsets

Five datasets with different similarity thresholds for matching articles are generated as described in Appendix A.1.3. Table A6 reports the point estimates and p-values from the comparison of emotionality and sentiment between online and offline headlines in each dataset. Figure A5 illustrates the distribution of headlines across the three sentiment categories for each subset.

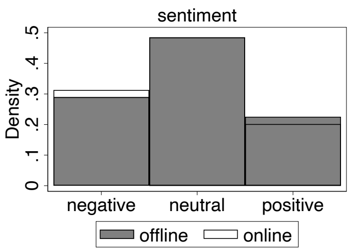
Table A6: Comparison of Emotionality and Sentiment between Online and Offline Headlines across Matched Datasets

Cosine Threshold	**Coefficient (Online)**	**Std. Error**	**p-value**	**Constant**
Outcome: Emotionality				
0.85	0.0129	0.0307	0.675	-0.0065
0.90	0.0154	0.0311	0.621	-0.0077
0.92	0.0092	0.0240	0.702	-0.0046
0.95	0.0060	0.0216	0.782	-0.0030
0.97	-0.0010	0.0254	0.970	0.0005
Outcome: Sentiment				
0.85	-0.0877	0.0311	0.005	0.0438
0.90	-0.0868	0.0321	0.007	0.0434
0.92	-0.0892	0.0347	0.010	0.0446
0.95	-0.0791	0.0205	0.000	0.0396
0.97	-0.0664	0.0313	0.034	0.0332
Observations	19,284 / 17,500 / 15,736 / 10,054 / 4,186			

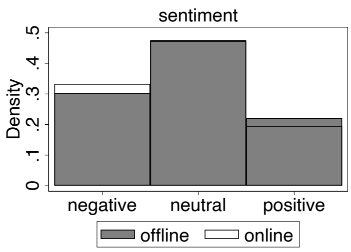
Notes: Each panel reports coefficients, bootstrapped standard errors clustered by newspaper (50 replications), p-values, and constants from regressions of headline emotionality or sentiment on an online dummy. A positive coefficient indicates higher values for online headlines relative to offline counterparts.

Figure A5: Robustness Check: Classifications with Matched Datasets

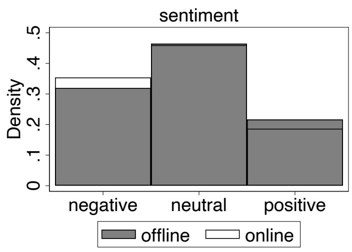
(a) Cosine Similarity 0.97



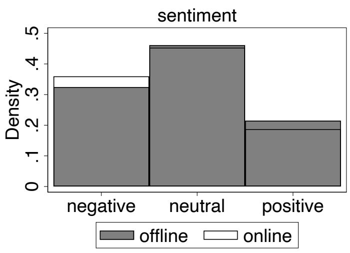
(b) Cosine Similarity 0.95



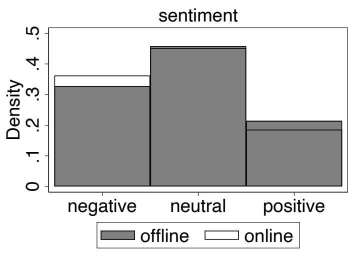
(c) Cosine Similarity 0.92



(d) Cosine Similarity 0.90



(e) Cosine Similarity 0.85



Notes: Figures A5(a) - A5(e) illustrate the distribution of sentiment classes in each of the matched subsets of the data. The offline headlines are shaded in gray.

A.1.6.4 Classification Methodology

Section A.1.1.3 shows that the fine-tuned GPT model performs best among all considered approaches for classifying headlines in the dataset. At the same time, all classification methods exhibit explanatory value,[7] which allows them to be used to replicate the analyses and to assess the extent to which the results depend on the chosen classification methodology.

This robustness check confirms that the finding that online headlines are, on average, more emotional holds across all classification algorithms considered. Based on these estimations, online headlines are between 0.20 and 0.45 standard deviations more emotional than offline headlines. When all available controls are included, as specified in equation A.4, the difference ranges from 0.14 to 0.27 standard deviations.

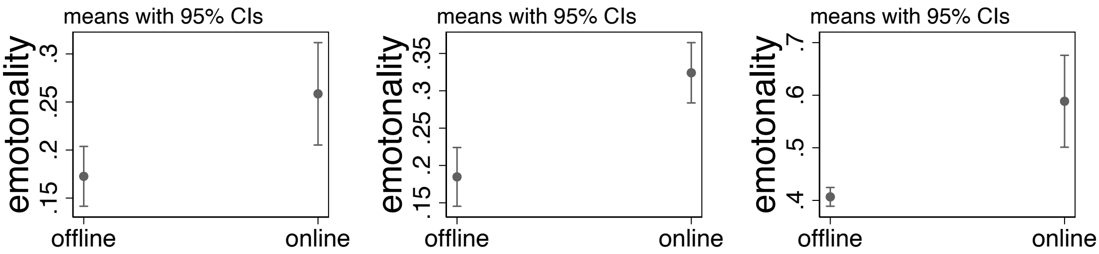
The result that online headlines are, on average, more negative is reproduced across all classification algorithms except VADER. Using VADER, the point estimates for sentiment are very small and positive. Without controls, the estimated differences in sentiment range from online headlines being 0.02 standard deviations more positive to 0.23 standard deviations more negative than offline ones. When all control variables are included, the difference ranges from 0.02 standard deviations more positive to 0.15 standard deviations more negative for online headlines.

Figure A6 displays the emotionality means and sentiment distributions obtained from the respective classification methods.

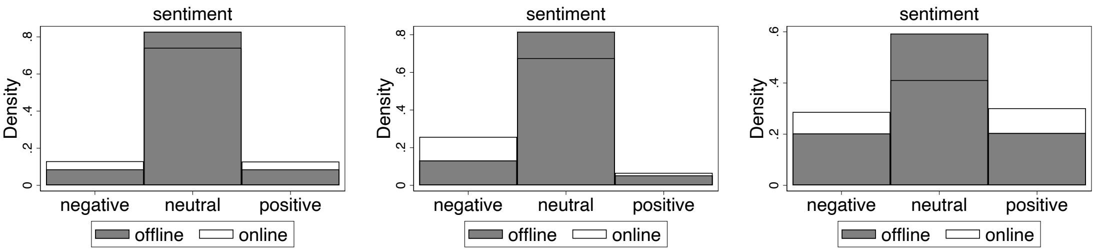
[7] That is, they perform better than random guessing.

Figure A6: Robustness Check: Classifications with Different Algorithms

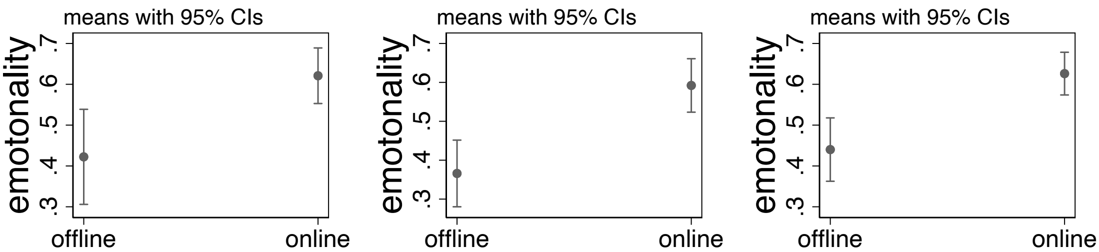
(a) SentiWS **(b)** LM **(c)** VADER



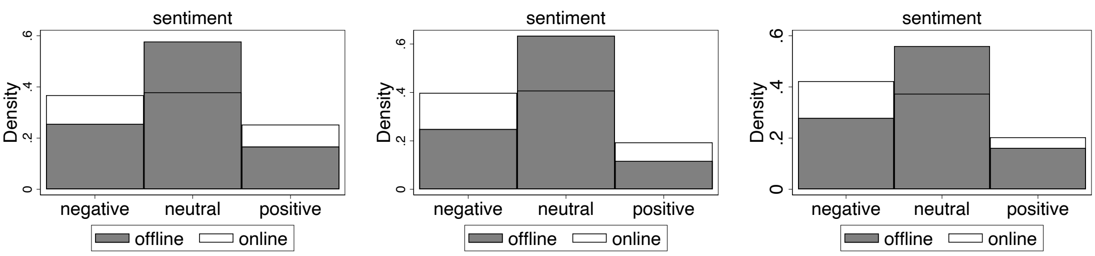
(d) SentiWS **(e)** LM **(f)** VADER



(g) pre-trained roBERTa **(h)** fine-tuned roBERTa **(i)** pre-trained GPT



(j) pre-trained roBERTa **(k)** fine-tuned roBERTa **(l)** pre-trained GPT



Notes: Figures A6(a) - A6(c) and A6(g) - A6(i) illustrate the means and 95 percent confidence intervals in emotionality online and offline (without controls, standard errors are bootstrapped as in the main specification). Figures A6(d) - A6(f) and A6(j) - A6(l) depict the distribution of the sentiment variable with the different classification algorithms. The offline headlines are shaded in gray.

A.1.6.5 News Markets

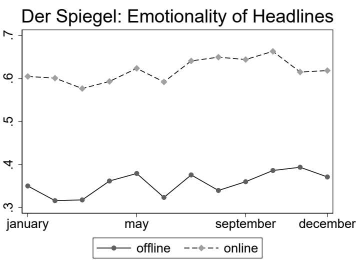
The main dataset consists only of articles on economic issues from German news outlets. Do the findings reproduce when articles on other topics or outlets from other countries are considered? To assess this I use additional datasets as described in the lower part of table 2.1. I consider headlines on all kinds of topics from *Der Spiegel*, *BILD* and *The New York Times* (but having a mostly much shorter time-frame available). As the content of the articles in not available in this data it is impossible to control for the article tonality in the subsequent regressions. Apart from this all estimations are equivalent to those described in section 2.2.2.2.

Comparing the estimates for *Der Spiegel* and *BILD* can help to understand how specific the findings are to articles on economic issues. The findings that headlines online are more emotional and negative than offline reproduce in these samples. Interestingly, the estimates are even bigger when compared to the estimate for the subset of economic articles for the same outlet. For example, online headlines from *Der Spiegel* are 0.52 standard deviations more emotional and 0.18 standard deviations more negative than offline (compared to 0.43 and 0.03 standard deviations in the economic article subset). A similar difference in effect size exists for *BILD*. This suggests that the estimations of the difference in sentiment of economic articles might be a relatively conservative estimate of the average difference for all topics.

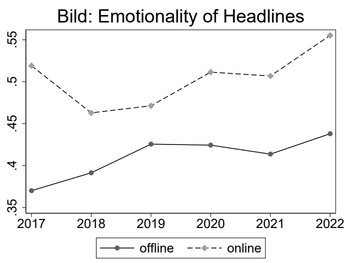
The estimates might however be specific to the German news market. I therefore collect data on headlines of one of the most important American newspapers, the *New York Times*, and repeat my analysis with it. Again, the main findings so far reproduce: Headlines in the *New York Times* are online on average more emotional and negative than offline. The effect size for emotionality is with 0.08 standard deviations however substantially smaller than the average for the German market (0.45 standard deviations). Figure A7 provides an overview over these comparisons by illustrating the emotionality of headlines online and offline in these additional datasets.

Figure A7: Robustness Check: Classifications with Different Datasets

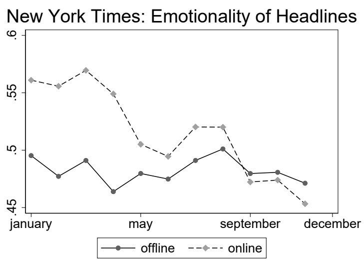
(a) Der Spiegel



(b) BILD



(c) The New York Times



Notes: Figures A7(a) - A7(c) illustrate the development of emotionality over time in the robustness datasets. The data from *Der Spiegel* and *The New York Times* are from 2021. Values for the offline headlines are illustrated by a full line. The online values are depicted with a dotted line.

A.1.7 Additional Descriptive Analyses on Mechanisms

A.1.7.1 Descriptive Evidence on Headline Competition in Offline Markets

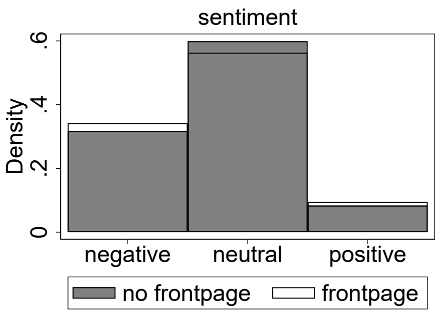
Does higher competition drive incentives to select emotional headlines? This section serves as a consistency check: If being published on the frontpage is correlated with higher emotionality of a headline, this makes the the hypothesis that the degree of competition is related to selecting emotional headlines more plausible.

The offline data of the main dataset does not come with information of the specific page a article was published on. This information is available in the offline part of the robustness datasets, therefore they are being used here. This means I can compare frontpage to non-frontpage headlines for *The New York Times*, *Der Spiegel* and *BILD*. The empirical strategy is identical to the one described in equations 2.1 to A.4, but instead of an online dummy a dummy for a headline being on the frontpage or not is being used. Additionally - just as in the robustness section - as these datasets do not contain the article content, I cannot control for content tonality. Note that any evidence presented here is again correlational, and not necessarily causal.

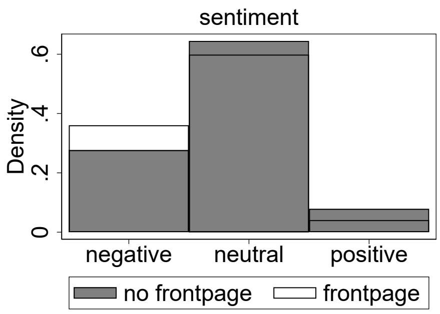
The comparison shows that frontpage headlines are indeed more emotional and more negative than headlines of stories in other parts of the papers. For *The New York Times*, frontpage headlines are on average 0.26 standard deviations more emotional and 0.18 standard deviations more negative than headlines on other pages. For *Der Spiegel*, frontpage headlines are on average 0.09 standard deviations more emotional and 0.18 standard deviations more negative than non-frontpage headlines. For *BILD*, frontpage headlines are on average 0.07 standard deviations more emotional and 0.02 standard deviations more negative than headlines in other parts of the newspaper. Figure A8 illustrates the distribution of the headline classifications for all three datasets. Regression results are available in Tables A7 - A12.

Figure A8: Sentiment of Frontpage- and Non-Frontpage Headlines

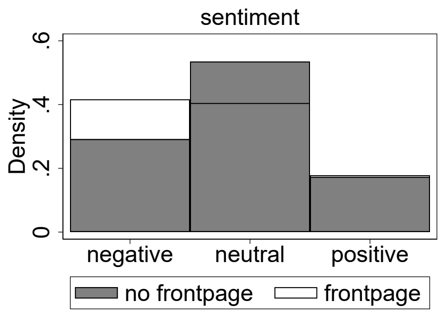
(a) BILD



(b) Der Spiegel



(c) The New York Times



Notes: Figures A8(a) - A8(c) illustrate the distribution of the sentiment variable for the offline data in the different datasets. Headlines that were not published on the frontpage are shaded in gray.

Table A7: OLS Estimates – Emotionality of BILD Headlines

	(1)	(2)	(3)	(4)
front-page	0.0725***	0.0227***	0.0706***	0.0154*
	(0.0084)	(0.0086)	(0.0084)	(0.0087)
article length				-0.0368**
				(0.0133)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	-0.0972***	-0.0472***	-0.1418***	-0.1264***
	(0.0034)	(0.0039)	(0.0305)	(0.0305)
R^2	0.0008	0.0077	0.0038	0.0124
Observations	97,576	97,576	97,576	97,576

Notes: Table A7 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table A8: OLS Estimates – Sentiment of BILD Headlines

	(1)	(2)	(3)	(4)
front-page	-0.0211**	0.0270***	-0.0207**	0.0339***
	(0.0083)	(0.0086)	(0.0083)	(0.0085)
article length				-0.0194
				(0.0122)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	0.0420***	-0.0064*	0.1094***	0.1043***
	(0.0033)	(0.0038)	(0.0296)	(0.0297)
R^2	0.0001	0.0069	0.0020	0.0099
Observations	97,576	97,576	97,576	97,576

Notes: Table A8 reports OLS estimates with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table A9: OLS Estimates – Emotionality of Spiegel Headlines

	(1)	(2)	(3)	(4)
front-page	0.0929	0.0628	0.0861	0.0788
	(0.0827)	(0.0846)	(0.0820)	(0.0892)
article length				-0.0106
				(0.0131)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	-0.4178***	-0.5526***	-0.4297***	-0.5401***
	(0.0140)	(0.0359)	(0.0433)	(0.0537)
R^2	0.0003	0.0347	0.0030	0.0372
Observations	4,896	4,896	4,896	4,896

Notes: Table A9 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table A10: OLS Estimates – Sentiment of Spiegel Headlines

	(1)	(2)	(3)	(4)
front-page	-0.1778***	-0.1169*	-0.1663**	-0.0471
	(0.0668)	(0.0682)	(0.0664)	(0.0724)
article length				-0.0242**
				(0.0110)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	0.1535***	0.1540***	0.1943***	0.1965***
	(0.0118)	(0.0286)	(0.0371)	(0.0449)
R^2	0.0014	0.0241	0.0035	0.0273
Observations	4,896	4,896	4,896	4,896

Notes: Table A10 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table A11: OLS Estimates – Emotionality of NYT Headlines

	(1)	(2)	(3)	(4)
front-page	0.2602***	0.1351***	0.2603***	0.1201***
	(0.0165)	(0.0163)	(0.0165)	(0.0171)
article length				0.0231***
				(0.0082)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	-0.0542***	-0.7082***	-0.0292	-0.7083***
	(0.0062)	(0.0906)	(0.0185)	(0.0913)
R^2	0.0082	0.1233	0.0086	0.1242
Observations	29,216	29,216	29,216	29,216

Notes: Table A11 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

Table A12: OLS Estimates – Sentiment of NYT Headlines

	(1)	(2)	(3)	(4)
front-page	-0.1755***	-0.0993***	-0.1756***	-0.0668***
	(0.0178)	(0.0182)	(0.0178)	(0.0189)
article length				-0.0512***
				(0.0073)
topic FE	no	yes	no	yes
time FE	no	no	yes	yes
Constant	0.0613***	0.4368***	0.0489**	0.0489***
	(0.0062)	(0.0663)	(0.0186)	(0.0186)
R^2	0.0037	0.0749	0.0041	0.0765
Observations	29,216	29,216	29,216	29,216

Notes: Table A12 reports OLS estimates in standard deviations with robust standard errors in parentheses. The numbers in parentheses above in the first line of the table correspond to the respective regression equation from section A.1.2. The non-front-page headlines are always the reference group. Control variables are article length in 1000 words and fixed effects as dummies for the different topics and days. *** p<0.01, ** p<0.05, * p<0.1

A.1.7.2 Anecdotal Evidence on Headline Competition in Online Markets

A plausible reason for increased incentives to capture readers' attention in online markets is higher competition on the headline level, which in turn might translate into more emotional headlines. To explore whether the degree of competition can really be related to the tonality of headlines I copy the idea behind the identification strategy of Meyer et al., 2024 and leverage the exclusion of a group of news outlets from popular news aggregators due to a legal dispute in Germany.

Context The exogenous variation in competition that I use is an exclusion of one of the newspapers in my sample, namely *Die Welt*, from several news aggregators due to a legal dispute in 2014. News aggregators are websites that combine news content from several sources, typically summarize the articles' content in one paragraph and provide the link to the original. Especially in the early days of these aggregators, traditional news companies oftentimes regarded them as a threat. They were afraid that many readers would be satisfied with the small summaries and thus use the aggregators as a substitute to consumption of online news from traditional providers. In this context, the German government passed the "ancillary copyright for press publishers". This bill allows traditional media companies to claim royalty fees if their content is used by other companies. However, short excerpts of text were excepted from the regulation, which led to uncertainty about if the new legislation applied to the paragraphs that news aggregators provided. The German copyright collecting society *VGM* published a pricing schedule for the reuse of its members' original content and threatened to file lawsuits against news aggregators that refused to pay[8]. As a response, a group of popular news aggregators (for example *gmx.de*, *web.de*, and *t-online.de*) removed all articles from of *VGM* members from their platforms in August 2014. They continued to show content from non-members. Thus, competition on the headline level decreased for *VGM* members, but not for non-members. I use this removal of *Die Welt* to explore how changes in competition can translate into changes of tonality.

Empirical Strategy Note that I only have data on three news outlets for the relevant time frame[9], namely *Der Spiegel*, *Die Welt* and *Die Zeit*. Out of the three available outlets, *Die Welt* was a *VGM* member, while the others were not. With that data I conduct two comparisons. First, I regard the development of the online versions *Die Welt* in contrast to the online versions of the other newspapers before

[8] See this press article for a summary of the discussion.

[9] A detailed description of the availability of the data at different points in time is denoted in Table 2.1.

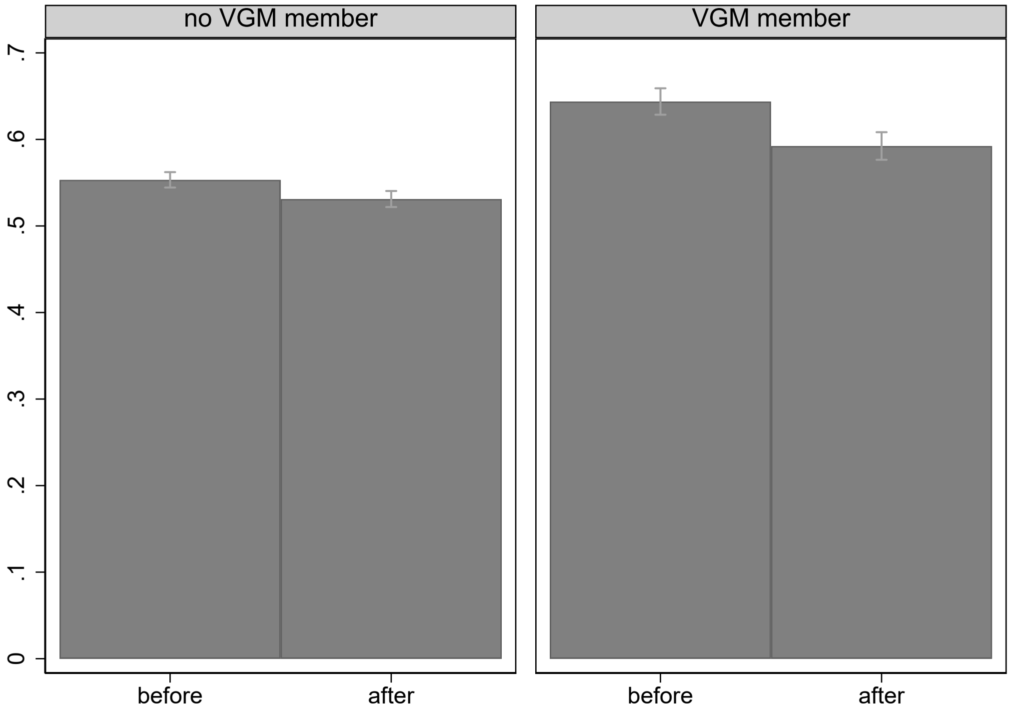
and after the removal. Second, I compare the tonality of the online headlines to the offline headlines of *Die Welt* before and after the removal.

The copyright bill was passed in March 2013 and the *VGM* members were removed from the aggregators in August 2014. As suggested in the empirical strategy of Meyer et al., 2024, I drop the time frame between these two dates to exclude potential confounders that may exist because of the ongoing debate about the legislation. Therefore, the regarded time frames are always a comparison of before March 2013 to after August 2014. In line with Meyer et al., 2024 I limit the time frame to 18 months before and after the removal, but also assess different time frames as a robustness checks later on. I graphically analyze the results and compute a back-of-the-envelope difference in differences estimator to get an idea about the effect size. Because of the low number of outlets in my sample (which hampers clustering standard errors on the outlet level), I don't run regressions to obtain this estimate.

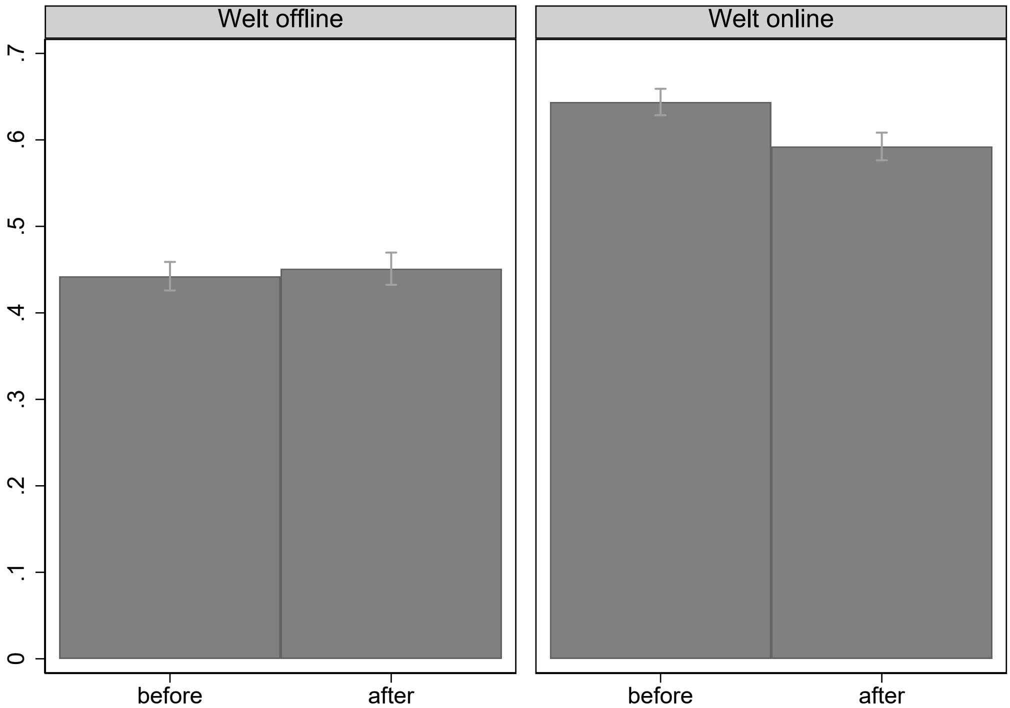
Main Finding The comparison reveals that the removal from the aggregators seems to have slightly reduced the share of emotional headlines at *Welt Online*. Figure A9 illustrates that comparison. Back-of-the-envelope estimations of the difference in differences suggest that the effect size is around 0.08 or 0.13 standard deviations (depending on which control group is used). I don't detect differences in terms of sentiment, as depicted in Figure A10.

Figure A9: Comparing the emotionality of headlines in the removed outlet to others

(a) Emotionality of the online headlines of the removed and non-removed outlets



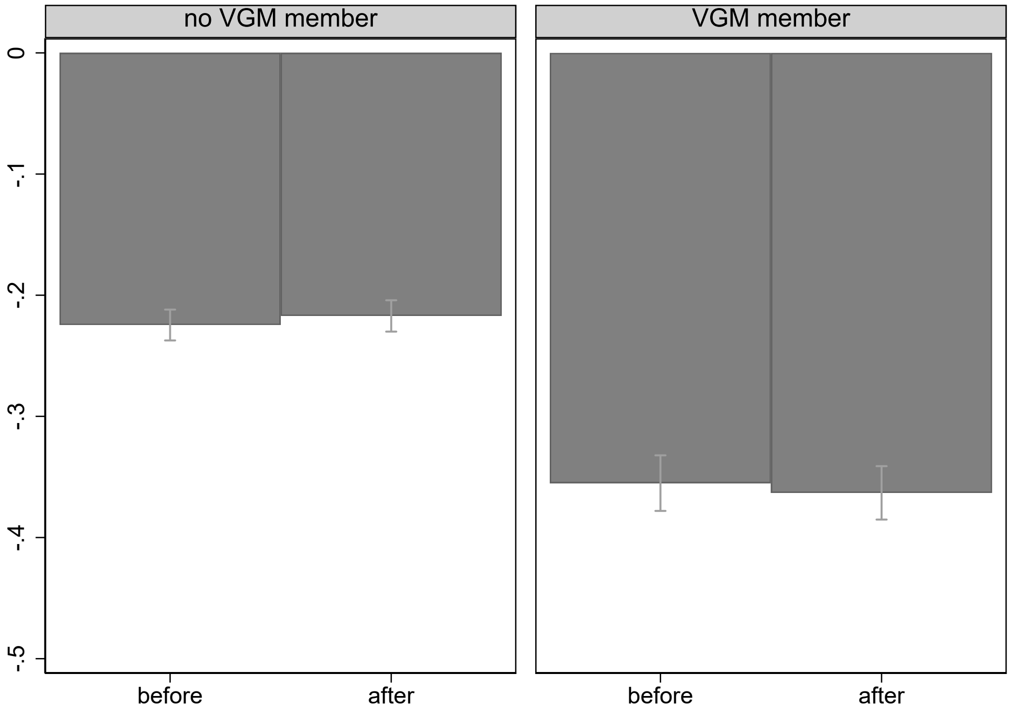
(b) Emotionality of the online and offline headlines of the removed outlet



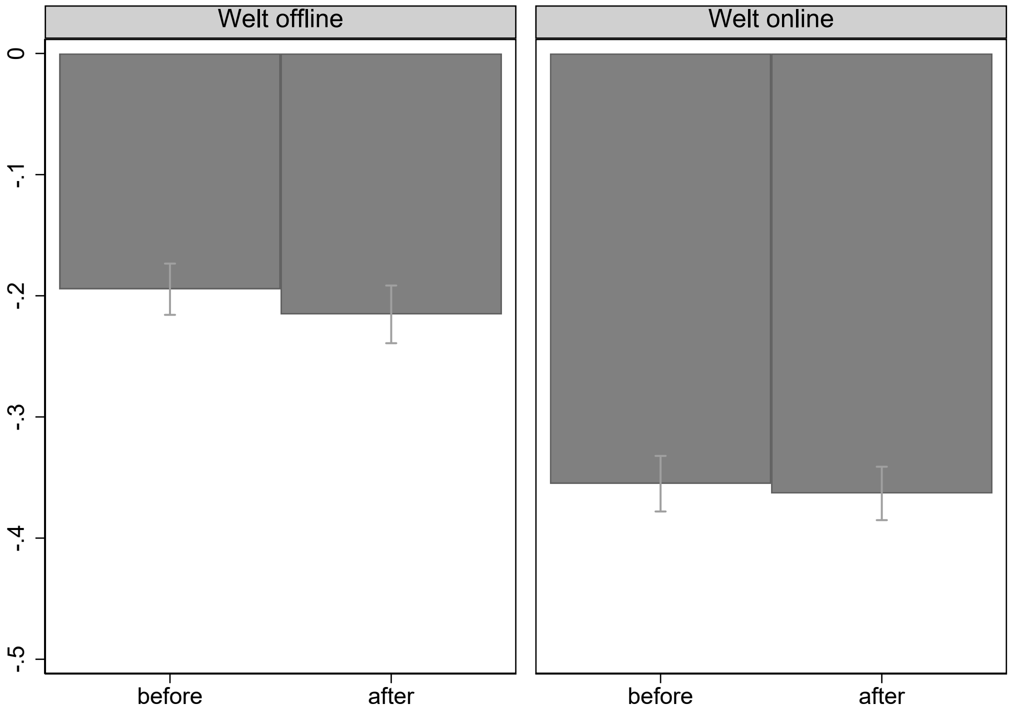
Notes: Figure A9(a) depicts the average emotionality of headlines at VGM members and non-members before and after the removal. Figure A9(b) illustrates the average emotionality of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Figure A10: Comparing the sentiment of headlines in the removed outlet to others

(a) Sentiment of the online headlines of the removed and non-removed outlets



(b) Sentiment of the online and offline headlines of the removed outlet



Notes: Figure A10(a) depicts the average sentiment of headlines at VGM members and non-members before and after the removal. Figure A10(b) illustrates the average sentiment of headlines at the online and offline versions of *Die Welt* before and after the removal. The time frame used for computing the averages is always 18 months before March 2013 and 18 months after August 2014.

Robustness: Alternative depended variables I assess the robustness of the finding by using the classifications of the other sentiment classifiers as described in Appendix A.1.1 as alternative outcomes. The finding that the removal reduces emotionality reproduces qualitatively with all other classification methodologies. The size of the estimated effect ranges from 0.05 to 0.17 standard deviations.

Robustness: Alternative time windows A concern could be that the result depends on the specific time frame used. To explore this, I repeat the comparison, but instead of a 18 months I change the time frame to 24 or 12 months before and after the removal. Also, I repeat the main analysis without removing the period from March 2013 to August 2014 from the dataset. These are the same robustness time frames that Meyer et al., 2024 test.

The result is robust to both an expansion and limitation of the regarded time frames. If the time frame is reduced to one year before and after the removal the estimate for the difference in differences is -0.06 standard deviations. When it is expanded to 2 years the resulting estimate is -0.08 standard deviations. Including the time frame between March 2013 to August 2014 in the analysis yields an estimate of -0.06 standard deviations.

Limitations As previously mentioned, the number of outlets used in this analysis is limited to three. Therefore, I recommend interpreting the results as first, suggestive evidence for the described mechanism. It seems to be a promising avenue for future research to further explore this relationship with more diverse samples.

A.1.7.3 Introduction of Paywalls

Figure A11: Paywall Introduction: *Der Spiegel*

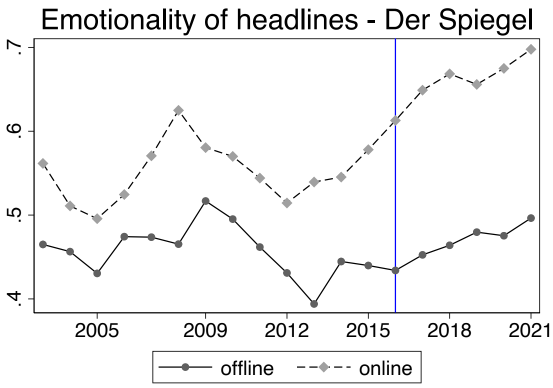


Figure A12: Paywall Introduction: *Die Welt*

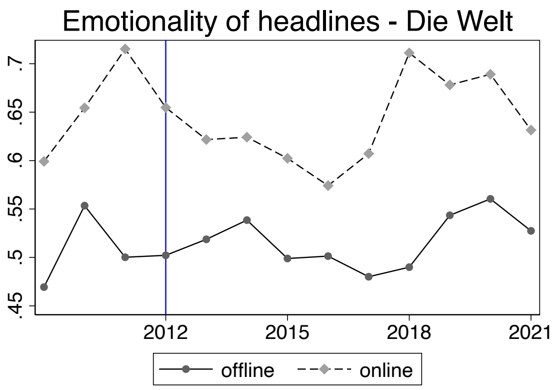
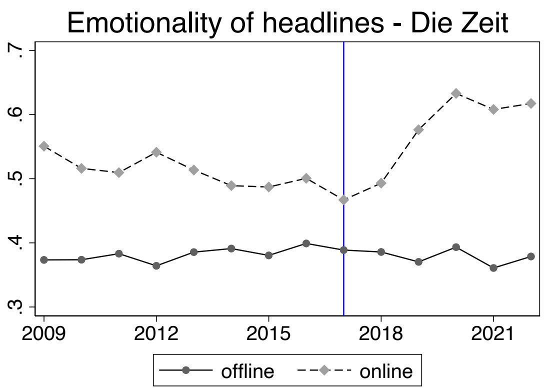


Figure A13: Paywall Introduction: *Die Zeit*



A.2 Supplementary Materials Causal Evidence from Journalists

A.2.1 Main Journalists' experiment

A.2.1.1 Article Pool

In total, six articles were available in the article pool. For each participating journalist, two articles were randomly drawn. Below, the full text and all associated headline options for each article are provided. The content was originally presented in German; here, an English translation is provided. Explanatory comments (which were not displayed to participants) are indicated by *blue, italic text*.

Article 1 – Politics

Main article text (displayed to journalists)

A group of Turkish producers has submitted an application to the European Commission to have the Döner Kebab legally protected as a "Guaranteed Traditional Specialty" (g.t.S.). If approved, Döner could only be sold under this name if it meets the criteria specified in the application, including certain ingredients and preparation methods such as a mild thyme flavor in the meat.

Additional paragraphs (displayed upon clicking "more"):

In Germany and Austria, however, there is opposition to the proposal. The German Hotel and Restaurant Association (Dehoga) criticizes that the preparation methods commonly used in Germany deviate from the Turkish specifications. Ingrid Hartges, chairwoman of the association, warns of "serious consequences" for gastronomy and consumers, as many existing Döner dishes would have to be renamed.

There are around 18,000 Döner shops in Germany that could be affected by the new regulation. The European Commission has set a three-month objection period during which affected parties can raise concerns. If sufficient and well-founded objections are submitted, the application could be rejected.

In addition, there are concerns that approval of the application could further increase Döner prices. Prices have already risen noticeably in recent years, and stricter requirements could drive them even higher.

Headline options shown to journalists

Existing headlines:
– Negative: *Döner shock: price explosion and standardized recipe looming*
– Neutral: *Will the EU soon regulate Döner?*
– Positive: *Quality legally defined? New EU law aims to protect Döner*

GPT-generated headlines:
– Negative: *EU protection for Döner threatens diversity and drives up prices*
– Neutral: *Application for EU protection of Döner sparks debate*
– Positive: *EU protection could ensure Döner quality across Europe*

Competing headline (shown only in competition condition)
"EU Commission plans to relax genetic engineering regulations"

Note: The content of the article belonging to the competing headline was used exclusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

The European Commission plans to relax regulations on genetic engineering in agriculture. After prior discussions, the Commission recently presented a draft law according to which foods based on genetically modified plants would no longer have to be labeled if the modifications could also occur naturally or through conventional breeding.

Specifically, the proposal concerns so-called "New Genomic Techniques" (NGT), which allow precise interventions in a plant's DNA. The new framework would also make it easier to research genetically modified food and feed. The Commission expects NGT methods to yield new crop varieties that can better adapt to climate change, require less water, or be more resistant to pests.

Moreover, new varieties could reach the market faster. Through the safe use of these techniques, farmers would gain access to more resilient plants that require fewer pesticides, said Commission Vice-President Frans Timmermans.

Stricter rules would continue to apply to plants whose genetic modifications cannot be achieved through conventional breeding, or where genes from unrelated organisms are transferred. These cases would remain subject to the existing restrictive framework. Organic and ecological farming would remain excluded from the use of NGT crops.

Article 2 – Technology

Main article text (displayed to journalists)

The International Monetary Fund (IMF) expects that artificial intelligence (AI) will affect around 40 percent of jobs worldwide. This is according to a recently published report by IMF Managing Director Kristalina Georgieva. She emphasized that the effects of AI could be both positive and negative, calling on governments to proactively take measures to mitigate potential risks.

Additional paragraphs (displayed upon clicking "more"):

In advanced economies such as the United States and Germany, around 30 percent of jobs could be negatively affected by AI, while another 30 percent could benefit from higher productivity, according to the report. Highly qualified workers, in particular, face both opportunities and challenges as certain tasks become increasingly automated.

In emerging and low-income countries, where about 40 percent of jobs could be affected, the lack of digital infrastructure may hinder the use of AI. This could exacerbate existing inequalities and make it harder to access the benefits of the technology.

The IMF recommends that governments expand social safety nets and offer retraining programs to help workers transition into new roles. In the long term, AI has the potential to increase global productivity if the right policy measures are implemented.

Georgieva stresses that the impact of AI will depend heavily on each country's economic situation and requires careful planning. One major challenge is ensuring that the benefits of AI are distributed fairly.

Headline options shown to journalists

Existing headlines:
– Negative: *Artificial intelligence: Almost 40 percent of jobs worldwide under threat*
– Neutral: *Artificial intelligence will transform the world of work*
– Positive: *IMF study: Germany among the best prepared for the AI revolution*

GPT-generated headlines:
– Negative: *Artificial intelligence threatens jobs, warns IMF*
– Neutral: *IMF: Artificial intelligence will transform global labor markets*
– Positive: *IMF sees great potential in artificial intelligence for labor markets worldwide*

Competing headline (shown only in competition condition)
"AI-generated images: How useful would a labeling requirement be?"

Note: The content of the article belonging to the competing headline was used exclusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

Technological progress in the field of artificial intelligence is advancing rapidly. Today, there are AI-generated images and videos that look deceptively real at first glance. As a result, they can also be misused for disinformation purposes.

Federal Minister of Justice Marco Buschmann (FDP) is therefore calling for a labeling requirement for AI-generated photos and texts. But how could such a rule be implemented? One option would be watermarks, which some providers already use, explains Xenia Klinge, a computational linguist and data scientist at the German Research Center for Artificial Intelligence. However, by law, such marking could only be required from domestic providers.

It is much more difficult, she adds, to label an AI-generated image retroactively. "We are almost in an arms race here—generation keeps improving, then detection improves, which again drives generation forward," says Klinge.

"We have reached a time when we can no longer trust all information," she warns. Photo and video manipulation have long existed, but the possibilities are now far easier. "We can no longer avoid thinking critically ourselves and verifying and contextualizing all information," the researcher concludes.

Article 3 – Science

Main article text (displayed to journalists)

The asteroid "Apophis" will pass Earth at a distance of about 30,000 kilometers on April 13, 2029 — closer than many satellites. With a diameter of roughly 350 meters, "Apophis" poses no danger to Earth. However, the opportunity to observe such a large asteroid from close range is extremely rare. Experts estimate that an event of this kind occurs only once every 5,000 to 10,000 years.

Additional paragraphs (displayed upon clicking "more"):

The European Space Agency (ESA) is planning the "Ramses" mission, a space probe that will accompany the asteroid during its flyby. The goal is to observe the effects of Earth's gravitational pull on "Apophis," which could cause changes to its surface. The mission is scheduled to launch in April 2028.

On April 13, 2029, about two billion people across Europe, Africa, and parts of Asia will be able to observe the flyby of "Apophis" with the naked eye. The asteroid will appear as a bright point in the night sky, weather permitting.

Although "Apophis" poses no immediate threat to Earth, the ESA emphasizes that the data collected could help prepare humanity for future asteroid threats. An impact from "Apophis" is considered excluded for at least the next 100 years.

Headline options shown to journalists

Existing headlines:
– Negative: *Mission to defend Earth: ESA sends satellite to "killer asteroid"*
– Neutral: *Asteroid mission: ESA plans "Ramses" probe to visit Apophis*
– Positive: *Asteroid "Apophis" to pass very close to Earth – "A great natural experiment"*

GPT-generated headlines:
– Negative: *Asteroid Apophis to come dangerously close to Earth in 2029*

– Neutral: *Asteroid Apophis to pass very near Earth in 2029*
– Positive: *Historic flyby of asteroid Apophis offers unique research opportunity*

Competing headline (shown only in competition condition)
"Amateur astronauts complete short trip to space"

Note: The content of the article belonging to the competing headline was used exclusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

At about 740 kilometers above Earth, the crew of the private "Polaris Dawn" mission completed their risky short trip into space. Live images from SpaceX showed mission commander Jared Isaacman and Sarah Gillis stretching out of the hatch of the "Crew Dragon" spacecraft in their space suits for several minutes. The hatch was then closed again, and the cabin pressure was restored.

SpaceX described the event as the "first commercial spacewalk." However, the amateur astronauts did not float freely in space; instead, they remained on a kind of ladder at the spacecraft's hatch. The exercise tested suit mobility, and numerous data points were collected.

According to former astronaut Ulrich Walter, the spacewalk phase was the most dangerous part of the entire mission. The exit was originally scheduled to begin earlier, at 8:23 a.m. CEST, but SpaceX did not specify the reason for the delay.

Unlike the International Space Station (ISS), the "Crew Dragon" has no airlock for extravehicular activity. Therefore, all four private astronauts on board had to put on their space suits, as they too were exposed to the vacuum of space and there was no breathable air left in the cabin.

Article 4 – Economics

Main article text (displayed to journalists)

The inflation rate in the euro area fell to 2.2 percent in August, according to recent data from the European statistical office Eurostat. In July, the rate of price increase had still been 2.6 percent. The inflation rate is thus approaching the European Central Bank's (ECB) target value of just under two percent.

Additional paragraphs (displayed upon clicking "more"):

In Germany, the inflation rate fell below the two-percent mark for the first time since March 2021, reaching 1.9 percent in August. This was mainly due to declining energy prices, which were 5.1 percent lower than a year earlier.

The so-called core inflation, which excludes volatile components such as energy and food, remained at 2.8 percent in the euro area in August. Despite the decline, ECB representatives warned against premature action and emphasized that price stability had not yet been permanently secured.

Despite the general decline in inflation, price pressures persist in certain sectors. Service prices, in particular, continued to rise above average in August, for example for insurance and hospitality services. Experts warn that this could make stabilizing inflation more difficult in the long term and require a cautious monetary policy approach.

Headline options shown to journalists

Existing headlines:
– Negative: *ECB target in sight — but central bankers urge caution*
– Neutral: *Inflation in the euro area falls to 2.2 percent*
– Positive: *Inflation near ECB target — path cleared for interest rate cuts*

GPT-generated headlines:
– Negative: *ECB warns of unstable inflation despite decline in the euro area*
– Neutral: *Inflation in the euro area falls to 2.2 percent*
– Positive: *Falling inflation in the euro area creates room for interest rate cuts*

Competing headline (shown only in competition condition)
"Statisticians publish new data on economy and public finances"

Note: The content of the article belonging to the competing headline was used ex-

clusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

Weak industrial production, a crisis in the construction sector, and reduced consumer spending due to inflation: the German economy remains in crisis and hovers on the brink of recession. In August, the Federal Statistical Office released details on how gross domestic product (GDP) developed in the second quarter, along with data on public finances for the first half of the year. Borrowing and compliance with the national debt brake have been major points of contention within the federal government for months.

In an initial estimate at the end of July, the Federal Statistical Office reported that the economy had contracted by 0.1 percent compared to the previous quarter. Economists were surprised by this result, as they had expected at least a slight recovery after a modest 0.2-percent growth in the first quarter.

As a result, Germany risks sliding back into recession. Economists speak of a "technical recession" when GDP shrinks in two consecutive quarters. The country's economic output had already declined slightly in 2023.

Most recently, business sentiment in Germany has continued to deteriorate. The business climate index calculated by the ifo Institute fell by 0.4 points in August to 86.6, marking the third consecutive drop in Germany's most important economic indicator.

Article 5 – Crime

Main article text (displayed to journalists)

The South East Hesse Police Headquarters has released the crime statistics for 2023. According to the report, the number of recorded offenses increased by around 5 percent to just under 43,500 cases. Despite this rise, South East Hesse remains the third-safest region in the state of Hesse relative to its population. In particular, the Main-Kinzig district (excluding Hanau) recorded a 1.7 percent decline to around 10,000 cases.

Additional paragraphs (displayed upon clicking "more"):

The statistics show a sharp increase in politically motivated crimes, which rose by 20 percent to 368 cases. Almost half of these offenses had a right-wing extremist background, while left-wing extremist offenses accounted for only 3.5 percent. A particularly notable rise was observed in antisemitic crimes.

The number of ATM explosions more than doubled, from four cases in 2022 to nine in 2023. The police take these offenses very seriously and, in each case, examine the possibility of charges for attempted homicide. The escape of perpetrators is also considered particularly dangerous.

Theft-related crimes increased by 12 percent, with self-service supermarkets especially affected. There, the number of cases rose by almost 30 percent to over 750 reported incidents. According to the police, this increase is partly linked to the declining standard of living among the population.

Headline options shown to journalists

Existing headlines:
– Negative: *Crime statistics: sharp rise in shoplifting cases*
– Neutral: *Crime in the Main-Kinzig district: police present 2023 statistics*
– Positive: *Third-safest region in Hesse: South East Hesse police present crime report*

GPT-generated headlines:
– Negative: *Rising crime: South East Hesse becoming less safe*
– Neutral: *Crime statistics 2023: new figures for South East Hesse*
– Positive: *South East Hesse remains one of the safest regions*

Competing headline (shown only in competition condition)
"Hesse calls for data retention to fight crime"

Note: The content of the article belonging to the competing headline was used exclusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

The state of Hesse aims to achieve the one-month retention of internet data to combat crime through an initiative in the Bundesrat. Minister President Boris Rhein (CDU) presented the proposal in Frankfurt on Friday. The plan calls for the blanket storage of IP addresses, which, according to Rhein, are essential for solving many serious crimes. He criticized the “quick freeze” procedure favored by the federal coalition government in Berlin as completely inadequate.

Hesse plans to introduce the proposal in the Bundesrat next Friday. Rhein said he saw good chances that other states would join the initiative, creating pressure to win support from within the federal government as well.

In a recent compromise on the use of communication data for criminal investigations, the federal coalition agreed on the “quick freeze” procedure. Under this approach, data are stored only once there is suspicion of a serious crime.

The European Court of Justice, in a ruling issued in September 2022, set strict limits on the retention of telecommunications data for law enforcement purposes in Germany. The judges ruled that the suspended German regulation on data retention was incompatible with EU law. However, they also stated that the retention of IP addresses could be permissible under certain conditions to combat serious crime.

Article 6 – Consumption

Main article text (displayed to journalists)

The consumer organization Stiftung Warentest tested 24 child car seats and baby carriers, finding considerable differences in quality. While some models achieved excellent results, four seats failed the test with a rating of “poor.” The products were evaluated for crash safety, handling, ergonomics, and pollutant content, with crash test results being particularly decisive.

Additional paragraphs (displayed upon clicking “more”):

The baby carrier “Nuna Pipa Urbn” received the highest overall rating of 1.6 (“good”). The “Cybex Cloud G i-Size” also performed well, earning a score of 1.8 (“good”). Notably, some lower-priced models also impressed. The “Avionaut Cosmo,” priced at 185 euros, achieved an overall score of 2.0 (“good”) and a crash safety score of 1.3

("very good").

Four child seats were downgraded due to high pollutant levels. In the case of the "Peg Perego Viaggio," the support leg of the Isofix base failed during the crash test, causing the seat to detach from the base. The German Automobile Club (ADAC) had also previously warned against this model.

Overall, 13 of the 24 tested child seats received a "good" rating, while four models were rated "poor" due to safety and pollutant issues.

Headline options shown to journalists

Existing headlines:
– Negative: *Consumer warning: "Danger to life" — never use THIS child seat*
– Neutral: *Stiftung Warentest tests 24 child seats*
– Positive: *These child car seats prove safe in consumer tests*

GPT-generated headlines:
– Negative: *High pollutant levels: four child seats fail safety tests*
– Neutral: *Stiftung Warentest: 24 child seats tested — major differences found*
– Positive: *Child seat test: affordable models impress with high safety standards*

Competing headline (shown only in competition condition)
"Turn it up: Headphones for children put to the test"

Note: The content of the article belonging to the competing headline was used exclusively in the readers' experiment and was not visible to journalists during this study.

Text of competing article (not displayed to journalists)

Most children's headphones look playfully colorful, but Stiftung Warentest's latest review paints a critical picture of the models currently available on the market. The study ("test," issue 5/2024) found that many children's headphones are too loud, offer poor sound quality — and one model was even found to be hazardous to health.

The testers examined a total of 12 children's headphones. Only one model, the

"Tigerbuddies" by Tigermedia, received an overall rating of "good," thanks to its sound quality and wearing comfort. It costs around 52 euros and was also the most expensive product in the test. The remaining models were priced between 18 and about 43 euros.

Of the 12 tested models, 7 received a "satisfactory" rating, one was rated "adequate," and 3 were rated "poor." Several failed to maintain the widely advertised 85-decibel volume limit. In addition, one headphone cable was found to contain pollutant levels exceeding EU limits.

Children's headphones differ from adult models mainly in band size and fit. Experts recommend that parents ensure children do not wear headphones for too long and keep listening volumes at safe levels.

A.2.1.2 Balance Table

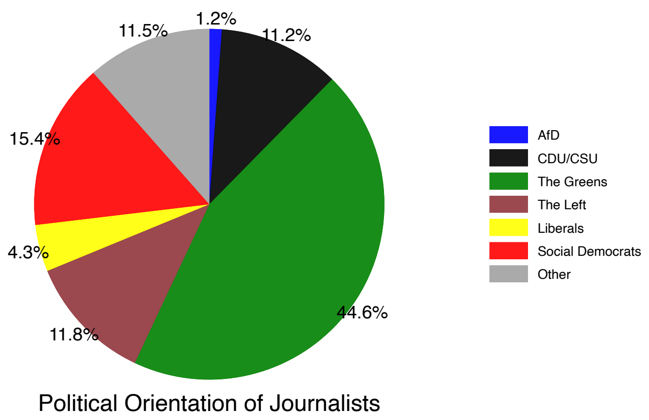
Table A13: Balance Table of Journalists (Participant Level)

	Flat pay (1)	Pay-per-click (2)	Pay-per-subscription (3)	t-test (4)	t-test (5)
Age	33.658	36.099	34.181	-2.441	-0.523
	(13.138)	(14.507)	(13.478)		
Male	43.2%	46.6%	45.7%	-3.3%	-2.5%
Editor	61.3%	63.4%	56.9%	-2.1%	4.4%
Region					
Outside Germany	5.2%	3.8%	1.7%	1.3%	3.4%
Eastern Germany	10.3%	13.7%	14.7%	-3.4%	-4.3%
Western Germany	81.3%	80.2%	82.8%	1.1%	-1.5%
No answer	3.2%	2.3%	0.9%	0.9%	2.4%
Political preference					
AfD	1.3%	0.0%	1.7%	1.3%	-0.4%
CDU/CSU	7.1%	9.9%	12.9%	-2.8%	-5.8%
The Greens	38.7%	36.6%	40.5%	2.1%	-1.8%
The Left	14.2%	6.1%	9.5%	8.1%**	4.7%
FDP	3.2%	4.6%	3.4%	-1.4%	-0.2%
No answer	5.8%	3.8%	0.0%	2.0%	5.8%***
SPD	9.7%	14.5%	11.2%	-4.8%	-1.5%
Other	7.7%	14.5%	10.3%	-6.8%*	-2.6%
Would not vote	10.3%	7.6%	10.3%	2.7%	0.0%
No response	1.9%	2.3%	0.0%	-0.4%	1.9%
Type of outlet					
Other online media	10.3%	10.7%	11.2%	-0.4%	-0.9%
No answer	0.0%	0.8%	1.7%	-0.8%	-1.7%
Online newspaper	27.1%	32.1%	32.8%	-5.0%	-5.7%
Print newspaper	34.2%	23.7%	31.0%	10.5%*	3.2%
Private broadcasting	7.1%	6.9%	5.2%	0.2%	1.9%
Other	3.2%	6.1%	5.2%	-2.9%	-1.9%
Public broadcasting	18.1%	19.8%	12.9%	-1.8%	5.1%
Education					
High school diploma	28.4%	26.0%	25.0%	2.4%	3.4%
Lower secondary school	1.3%	0.8%	0.9%	0.5%	0.4%
No answer	1.3%	2.3%	0.0%	-1.0%	1.3%
Intermediate school	2.6%	2.3%	1.7%	0.3%	0.9%
University	66.5%	68.7%	72.4%	-2.3%	-6.0%
Participants	155	131	116		

Notes: Column (4) shows flat minus pay-per-click; Column (5) shows flat minus pay-per-subscription. Analysis at the participant level (unique ResponseId). For continuous variables, the second line reports standard deviations in parentheses. For binary and categorical variables, entries are percentages. Stars indicate pairwise test significance vs. control: ***, **, * denote $p < 0.01$, $p < 0.05$, and $p < 0.10$, respectively. Variables: *Region*, *Political preference*, *Type of outlet*, *Education*, *Editor*, *Male*, *Age*.

A.2.1.3 Sample Descriptives

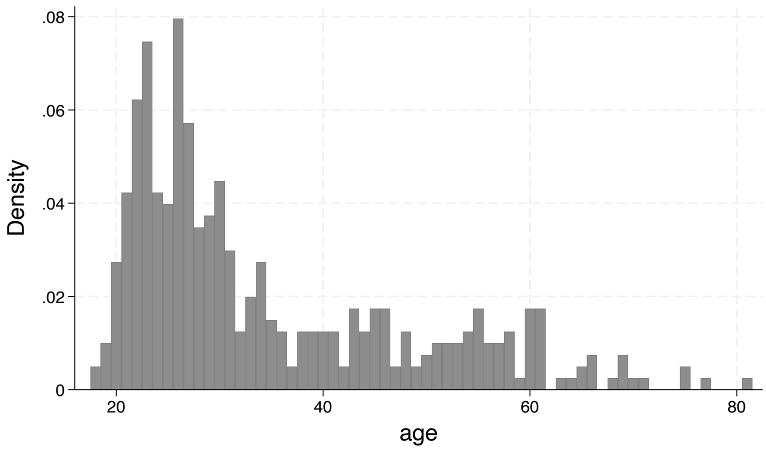
Figure A14: Distribution of Political Preferences in the Journalist Sample



Notes: Figure A14 gives an overview of the distribution of political preferences in the journalist sample.

Compared to the overall German population, political preferences in the sample are skewed toward left-leaning parties, consistent with previous survey evidence among journalists.

Figure A15: Distribution of Age in the Journalist Sample



Notes: Figure A15 gives an overview of the distribution of age in the journalist sample.

A.2.1.4 Results

Figure A16: Most frequently mentioned criteria for headline selection

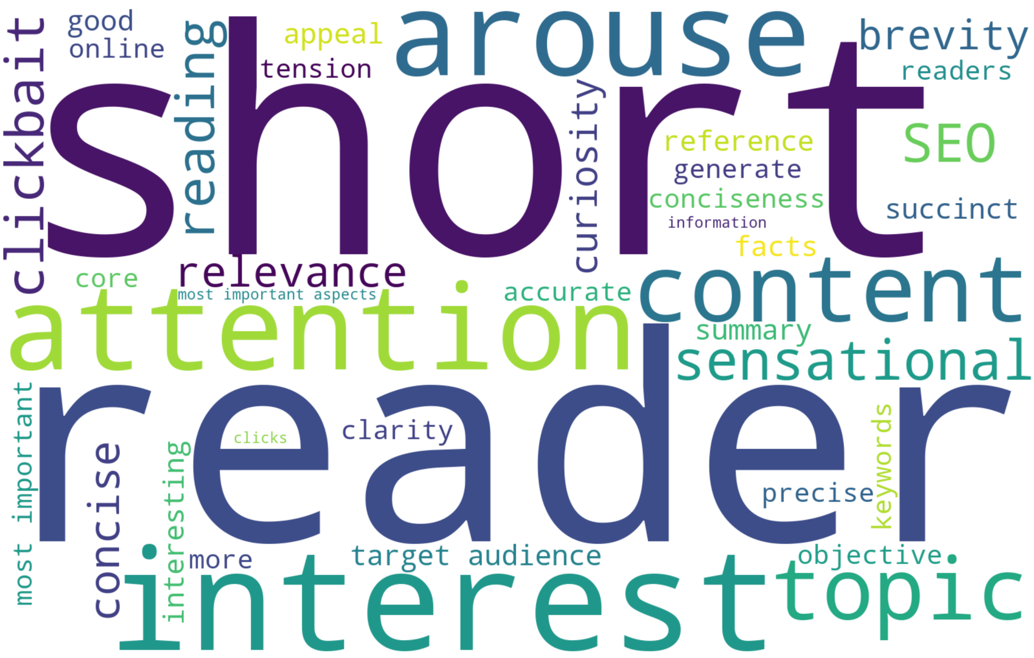


Table A14: OLS Estimates – ATE on Headline Choice

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	-0.1623***	-0.1706***	0.1109***	0.1220***
	(0.0445)	(0.0412)	(0.0251)	(0.0233)
Pay-per-abo	-0.1980***	-0.2092***	0.0737**	0.0737***
	(0.0483)	(0.0452)	(0.0267)	(0.0248)
age		0.0008		0.0002
		(0.0017)		(0.0009)
male		-0.0814**		-0.0120
		(0.0379)		(0.0209)
editor		-0.0330		-0.0083
		(0.0381)		(0.0206)
Region (ref.: other)				
East Germany		-0.1699		-0.0452
		(0.1201)		(0.0622)
West Germany		-0.2071*		0.0144
		(0.1111)		(0.0582)
No answer		-0.2788*		-0.0212
		(0.1642)		(0.0966)
Political orientation (ref.: AfD)				
CDU/CSU		0.4958***		-0.2289***
		(0.1356)		(0.0751)
The Greens		0.5731***		-0.2041***
		(0.1229)		(0.0711)
The Left		0.5185***		-0.2342***
		(0.1315)		(0.0790)
FDP		0.4428***		-0.2128**
		(0.1618)		(0.0846)
No answer		0.3152**		-0.1330*
		(0.1361)		(0.0784)
SPD		0.4534***		-0.2470***
		(0.1278)		(0.0751)
Other		0.5324***		-0.1872**
		(0.1298)		(0.0749)
Would not vote		0.5719***		-0.2434***
		(0.1281)		(0.0754)
Newsroom type (ref.: online)				
No answer		-0.1291		0.1267***
		(0.2003)		(0.0410)
Print		-0.1067*		0.0458
		(0.0620)		(0.0381)
Private broadcast		-0.0569		0.0695
		(0.0880)		(0.0475)
Public broadcast		-0.0816		0.0057
		(0.0626)		(0.0399)
Other		-0.0650		-0.0179
		(0.0876)		(0.0634)
Education (ref.: other)				
Lower secondary (Hauptschule)		-0.3877		0.0322
		(0.2928)		(0.1220)
No answer		0.2181*		-0.1754***
		(0.1220)		(0.0668)
Intermediate (Realschule)		0.4045***		-0.1365**
		(0.1199)		(0.0681)
University		0.0958**		-0.0469**
		(0.0451)		(0.0222)
Topic fixed effects	no	yes	no	yes
Constant	0.1265***	-0.2429	0.6013***	0.9201***
	(0.0257)	(0.1685)	(0.0171)	(0.1006)
R^2	0.0118	0.0719	0.0101	0.0947
Observations	4,020	4,020	4,020	4,020

Notes: Table A29 reports OLS estimates with robust standard errors clustered at the level of the individual journalist (402 clusters) in parentheses. The flat-rate (fixed pay) group serves as the reference category. The dependent variables are the sentiment and emotionality of journalists' headline choices. *** p<0.01, ** p<0.05, * p<0.1.

Table A15: ATE on Headline Choice for Free-Text Headlines

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	-0.1416**	-0.1397**	0.0694*	0.0880**
	(0.0651)	(0.0604)	(0.0407)	(0.0361)
Pay-per-abo	-0.1608**	-0.1445**	0.0520	0.0656*
	(0.0684)	(0.0603)	(0.0413)	(0.0369)
age		-0.0015		-0.0000
		(0.0021)		(0.0013)
male		-0.0693		0.0175
		(0.0537)		(0.0329)
editor		-0.0109		0.0110
		(0.0531)		(0.0339)
Region (ref.: other)				
East Germany		-0.1062		0.0390
		(0.1836)		(0.0820)
West Germany		-0.2588		0.0404
		(0.1692)		(0.0693)
No answer		-0.0436		-0.0814
		(0.2185)		(0.1503)
Political orientation (ref.: AfD)				
CDU/CSU		0.4759**		-0.2692**
		(0.2055)		(0.1138)
The Greens		0.6318***		-0.2915***
		(0.1937)		(0.1058)
The Left		0.5102**		-0.3141***
		(0.2114)		(0.1143)
FDP		0.4949**		-0.1946
		(0.2301)		(0.1386)
No answer		0.5695***		-0.2101*
		(0.1969)		(0.1085)
SPD		0.6572***		-0.2989***
		(0.2058)		(0.1127)
Other		0.4918**		-0.2388**
		(0.1973)		(0.1127)
Would not vote		0.5719***		-0.2430*
		(0.1281)		(0.1447)
Newsroom type (ref.: online)				
No answer		-0.1291		0.1210
		(0.2003)		(0.1142)
Online: daily/magazine		0.0332		0.0432
		(0.0889)		(0.0608)
Print: daily/magazine		-0.0591		0.0754
		(0.0899)		(0.0608)
Private broadcast		0.1206		0.0641
		(0.1225)		(0.0882)
Other		0.0470		0.0033
		(0.1324)		(0.0880)
Public broadcast		-0.1520		0.0610
		(0.0995)		(0.0642)
Education (ref.: other)				
Lower secondary (Hauptschule)		-0.3877		0.1143
		(0.2928)		(0.2528)
No answer		0.2181*		-0.0259
		(0.1220)		(0.1199)
Intermediate (Realschule)		0.4045***		-0.2555*
		(0.1199)		(0.1392)
University/FH		0.0958**		0.0165
		(0.0451)		(0.0361)
Topic fixed effects	no	yes	no	yes
Constant	-0.1065**	-0.5693**	0.6290***	1.0523***
	(0.0431)	(0.2756)	(0.0281)	(0.1497)
R^2	0.0087	0.2768	0.0042	0.3013
Observations	804	804	804	804

Notes: Table A15 reports OLS estimates for the subgroup of free-text headlines. Robust standard errors (in parentheses) are clustered at the journalist level (402 clusters). The flat-rate (fixed pay) group is the reference category. Columns (2) and (4) include controls: age, male, region, political orientation (ref.: AfD), newsroom type, editor status, education, and topic fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table A16: ATE on Headline Choice for Existing Headlines, by Competition

	No competition				With competition			
	Sentiment		Emotionality		Sentiment		Emotionality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pay-per-click	-0.1869***	-0.1984***	0.0859**	0.0955**	-0.1810**	-0.2099***	0.0868**	0.0994***
	(0.0716)	(0.0677)	(0.0385)	(0.0390)	(0.0705)	(0.0682)	(0.0384)	(0.0380)
Pay-per-abo	-0.2902***	-0.3075***	0.0596	0.0539	-0.1969**	-0.2262***	0.0672	0.0750*
	(0.0715)	(0.0670)	(0.0419)	(0.0415)	(0.0765)	(0.0750)	(0.0419)	(0.0414)
age		0.0004		-0.0036***		0.0023		0.0001
		(0.0026)		(0.0014)		(0.0026)		(0.0015)
male		-0.0527		0.0125		-0.0292		-0.0205
		(0.0593)		(0.0327)		(0.0612)		(0.0343)
editor		0.0172		-0.0001		-0.1070		-0.0070
		(0.0590)		(0.0338)		(0.0630)		(0.0337)
Region								
East Germany		-0.1026		-0.0343		-0.2392		-0.1808*
		(0.1900)		(0.1148)		(0.1740)		(0.1049)
West Germany		-0.0876		0.0335		-0.2722*		-0.1145
		(0.1752)		(0.1078)		(0.1564)		(0.0937)
No answer		-0.1682		-0.1147		-0.1429		-0.0832
		(0.2472)		(0.1753)		(0.2510)		(0.1373)
Pol. orientation								
CDU/CSU		0.8156***		-0.1952		0.5066**		-0.0188
		(0.1614)		(0.1205)		(0.2069)		(0.1176)
The Greens		0.8439***		-0.1481		0.5757***		0.0265
		(0.1374)		(0.1098)		(0.1827)		(0.1103)
The Left		0.7440***		-0.1840		0.3974**		-0.0413
		(0.1557)		(0.1223)		(0.2002)		(0.1243)
FDP		0.6168***		-0.0370		0.3951*		-0.1806
		(0.2091)		(0.1225)		(0.2281)		(0.1398)
SPD		0.6116***		-0.0923		0.3816**		-0.0334
		(0.1485)		(0.1150)		(0.1902)		(0.1152)
Other		0.6582***		-0.1184		0.6888***		0.0063
		(0.1675)		(0.1214)		(0.1941)		(0.1171)
Would not vote		0.6129***		-0.1495		0.5144***		-0.1169
		(0.1546)		(0.1197)		(0.1908)		(0.1180)
News type								
No answer		0.3995		0.1132*		-0.2185		0.2414***
		(0.3319)		(0.0684)		(0.5352)		(0.0750)
Online: newspaper		0.0479		0.0321		-0.1182		0.0874
		(0.0964)		(0.0606)		(0.1041)		(0.0616)
Print: newspaper		-0.0419		0.0088		-0.1766*		0.0376
		(0.0978)		(0.0625)		(0.1073)		(0.0628)
Private broadcast		-0.0015		-0.0232		-0.3063**		0.1040
		(0.1458)		(0.0851)		(0.1445)		(0.0786)
Other		-0.1306		-0.0252		-0.2027		-0.1236
		(0.1436)		(0.1030)		(0.1248)		(0.1069)
Public broadcast		0.0518		-0.0363		-0.2004*		0.0251
		(0.0996)		(0.0660)		(0.1101)		(0.0683)
Education								
Lower secondary		-0.5981		0.0176		-0.5251		0.0873
		(0.3793)		(0.2282)		(0.3782)		(0.1326)
No answer		0.1725		-0.4182**		-0.2127		-0.2843
		(0.2232)		(0.1765)		(0.2276)		(0.2530)
Intermediate		0.3551**		-0.1649		0.5975***		0.0825
		(0.1735)		(0.1266)		(0.1975)		(0.0924)
University/FH		0.0734		-0.0558		0.0865		-0.0588
		(0.0749)		(0.0383)		(0.0771)		(0.0368)
Topic FE	no	yes	no	yes	no	yes	no	yes
Constant	0.3548***	-0.4148*	0.6774***	1.0050***	0.1581***	-0.1431	0.6613***	0.8661***
	(0.0439)	(0.2321)	(0.0277)	(0.1668)	(0.0459)	(0.2841)	(0.0270)	(0.1598)
R^2	0.0218	0.1690	0.0069	0.1119	0.0119	0.1026	0.0072	0.0943
Observations	804	804	804	804	804	804	804	804

Notes: Table A16 reports OLS estimates for the subset of existing headlines, split by the existence of competition. Robust standard errors (in parentheses) are clustered at the journalist level (402 clusters). The flat-rate (fixed pay) group is the reference category. Columns (2), (4), (6), and (8) include controls: age, male, region, political orientation (ref.: AfD), newsroom type, editor status, education, and *topic fixed effects.* *** p<0.01, ** p<0.05, * p<0.1.

Table A17: ATE on Headline Choice for GPT-Generated Headlines, by Competition

	No competition				With competition			
	Sentiment		Emotionality		Sentiment		Emotionality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pay-per-click	-0.1258*	-0.1385**	0.2205***	0.2341***	-0.1763**	-0.2062***	0.0920**	0.0909**
	(0.0667)	(0.0638)	(0.0407)	(0.0398)	(0.0698)	(0.0702)	(0.0419)	(0.0411)
Pay-per-abo	-0.1129*	-0.1321*	0.1163**	0.1115**	-0.2293***	-0.2524***	0.0734*	0.0648
	(0.0678)	(0.0684)	(0.0435)	(0.0426)	(0.0737)	(0.0727)	(0.0440)	(0.0434)
age		0.0041*		0.0014		0.0007		0.0031**
		(0.0023)		(0.0014)		(0.0026)		(0.0015)
male		-0.1194**		-0.0147		-0.1309**		-0.0570
		(0.0561)		(0.0358)		(0.0636)		(0.0351)
editor		0.0101		-0.0220		-0.1096*		-0.0232
		(0.0579)		(0.0369)		(0.0635)		(0.0360)
Region								
East Germany		-0.2160		-0.0946		-0.1876		0.0500
		(0.1537)		(0.0889)		(0.1666)		(0.1139)
West Germany		-0.2242		-0.0231		-0.2360		0.1392
		(0.1387)		(0.0760)		(0.1471)		(0.1037)
No answer		-0.5914***		0.0295		-0.4577*		0.1443
		(0.2035)		(0.1496)		(0.2447)		(0.1365)
Pol. orientation								
CDU/CSU		0.2713		-0.3253		0.4469		-0.3352***
		(0.2076)		(0.1926)		(0.4235)		(0.1137)
The Greens		0.4223**		-0.2648		0.4364		-0.3411***
		(0.1900)		(0.1861)		(0.4156)		(0.1058)
The Left		0.5390**		-0.2379		0.4317		-0.3926***
		(0.2069)		(0.1911)		(0.4222)		(0.1181)
FDP		0.3227		-0.2250		0.4434		-0.4239***
		(0.2389)		(0.1990)		(0.4332)		(0.1414)
SPD		0.2933		-0.3654*		0.3724		-0.4177***
		(0.1906)		(0.1896)		(0.4176)		(0.1079)
Other		0.3523*		-0.2769		0.5865		-0.2850**
		(0.2074)		(0.1920)		(0.4242)		(0.1142)
Would not vote		0.4376**		-0.2944		0.5477		-0.3931***
		(0.1991)		(0.1880)		(0.4203)		(0.1119)
News type								
No answer		-0.4104		0.0678		-0.2397		0.1010
		(0.3704)		(0.1227)		(0.3709)		(0.1617)
Online: newspaper		-0.0760		0.0496		-0.0834		0.0827
		(0.0955)		(0.0600)		(0.1079)		(0.0608)
Print: newspaper		-0.1472		0.0791		-0.0978		0.0286
		(0.0940)		(0.0625)		(0.1103)		(0.0633)
Private broadcast		-0.0780		0.1157		-0.0338		0.0840
		(0.1439)		(0.0716)		(0.1579)		(0.0786)
Other		-0.0512		0.0646		-0.0513		-0.0096
		(0.1332)		(0.0862)		(0.1597)		(0.0981)
Public broadcast		-0.0999		-0.0105		-0.0184		-0.0083
		(0.0990)		(0.0667)		(0.1184)		(0.0713)
Education								
Lower secondary		-0.3653		-0.1937		-0.3638		0.1434
		(0.3247)		(0.1509)		(0.3195)		(0.1594)
No answer		0.3459		-0.1027		0.5010**		-0.0571
		(0.2246)		(0.1627)		(0.2219)		(0.1418)
Intermediate		0.3620**		-0.1629		0.3344*		-0.1796*
		(0.1445)		(0.1327)		(0.1962)		(0.1049)
University/FH		0.0779		-0.0613		0.1427**		-0.0774*
		(0.0689)		(0.0390)		(0.0722)		(0.0398)
Topic FE	no	yes	no	yes	no	yes	no	yes
Constant	0.1258***	-0.0224	0.4742***	0.8376***	0.1000**	-0.0909	0.5645***	0.8302***
	(0.0390)	(0.2456)	(0.0281)	(0.2123)	(0.0442)	(0.4389)	(0.0295)	(0.1582)
R^2	0.0059	0.1089	0.0354	0.1759	0.0163	0.0826	0.0072	0.1197
Observations	804	804	804	804	804	804	804	804

Notes: Table A17 reports OLS estimates for GPT-generated headlines, split by the existence of competition. Robust standard errors (in parentheses) are clustered at the journalist level (402 clusters). The flat-rate (fixed pay) group is the reference category. Columns (2), (4), (6), and (8) include controls: age, male, region, political orientation (ref.: AfD), newsroom type, editor status, education, and *topic fixed effects.* *** p<0.01, ** p<0.05, * p<0.1.

Table A18: ATE on Emotionality – Heterogeneity

	Age (Old)		Male		Left		University	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pay-per-click	0.1104***	0.0984***	0.0896**	0.0914***	0.1853***	0.1853***	0.0852**	0.0962**
	(0.0308)	(0.0301)	(0.0343)	(0.0304)	(0.0409)	(0.0409)	(0.0386)	(0.0402)
PPC × Moderator	0.0067	0.0653	0.0500	0.0693	-0.0914*	-0.0914*	0.0556	0.0474
	(0.0517)	(0.0503)	(0.0500)	(0.0467)	(0.0521)	(0.0521)	(0.0524)	(0.0522)
Pay-per-abo	0.0479	0.0359	0.0374	0.0408	0.1637***	0.1637***	0.0601	0.0204
	(0.0340)	(0.0321)	(0.0363)	(0.0326)	(0.0439)	(0.0439)	(0.0470)	(0.0436)
PPA × Moderator	0.0730	0.1085**	0.0827	0.0742	-0.1430**	-0.1430**	0.0343	0.0822
	(0.0538)	(0.0506)	(0.0532)	(0.0491)	(0.0578)	(0.0578)	(0.0602)	(0.0557)
controls	no	yes	no	yes	no	yes	no	yes
Constant	0.6115***	0.9396***	0.6261***	0.9447***	0.6260***	0.6260***	0.6587***	0.9022***
	(0.0219)	(0.0899)	(0.0233)	(0.0946)	(0.0951)	(0.0951)	(0.0266)	(0.1040)
R^2	0.0112	0.0966	0.0118	0.0960	0.0967	0.0967	0.0171	0.0939
Observations	4,020	4,020	4,020	4,020	3,216	3,216	3,216	3,216

Notes: OLS with robust standard errors clustered at the journalist level (402 clusters). The moderator is *old* in columns (1)–(2), *male* in (3)–(4), *left* in (5)–(6), and *university* degree in (7)–(8). The omitted pay arm is flat rate. "Controls" include, where applicable, age, male, region, political orientation, newsroom type, editor status, and education. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A19: ATE on Sentiment – Heterogeneity

	Age (Old)		Male		Left		University	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pay-per-click	-0.1093**	-0.1357***	-0.0950	-0.1181**	-0.2139***	-0.2139***	-0.2135**	-0.2205**
	(0.0524)	(0.0517)	(0.0580)	(0.0546)	(0.0744)	(0.0744)	(0.0903)	(0.0897)
PPC × Moderator	-0.1321	-0.0806	-0.1440	-0.1193	0.0567	0.0567	0.0657	0.0521
	(0.0950)	(0.0876)	(0.0886)	(0.0835)	(0.0932)	(0.0932)	(0.1070)	(0.1051)
Pay-per-abo	-0.2076***	-0.2333***	-0.1148*	-0.1408**	-0.3548***	-0.3548***	-0.2133**	-0.2479**
	(0.0588)	(0.0565)	(0.0661)	(0.0611)	(0.0800)	(0.0800)	(0.0948)	(0.0970)
PPA × Moderator	0.0251	0.0650	-0.1816*	-0.1543*	0.2319**	0.2319**	0.0051	0.0236
	(0.1013)	(0.0909)	(0.0947)	(0.0910)	(0.1032)	(0.1032)	(0.1147)	(0.1133)
controls	no	yes	no	yes	no	yes	no	yes
Constant	0.1184***	-0.2354	0.1307***	-0.2858*	0.2872*	0.2872*	0.1587***	-0.0771
	(0.0305)	(0.1757)	(0.0354)	(0.1600)	(0.1488)	(0.1488)	(0.0548)	(0.2163)
R^2	0.0136	0.0729	0.0187	0.0735	0.0186	0.0829	0.0145	0.0777
Observations	4,020	4,020	4,020	4,020	3,216	3,216	3,216	3,216

Notes: OLS with robust standard errors clustered at the journalist level (402 clusters). Moderators as in Table A18. The omitted pay arm is flat rate. "Controls" include, where applicable, age, male, region, political orientation, newsroom type, editor status, and education. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A20: Within-Subjects Effects of Competition on Headline Choice

	Sentiment (1)	Emotionality (2)
Competition	-0.1219***	0.0112
	(0.0243)	(0.0147)
Constant	0.1312***	0.6511***
	(0.0121)	(0.0074)
R^2 (within)	0.008	0.000
Observations	3,216	3,216
Individuals	402	402

Notes: Fixed-effects regressions with robust standard errors clustered at the individual journalist level (402 clusters). The dependent variable is either sentiment or emotionality of the selected headline. *** p<0.01, ** p<0.05, * p<0.1.

Table A21: Competition × Incentive Scheme: Headline Sentiment and Emotionality

	Sentiment (1)	Emotionality (2)
Competition	-0.1113***	0.0371
	(0.0397)	(0.0256)
Competition × Pay-per-click	-0.0223	-0.0638*
	(0.0589)	(0.0366)
Competition × Pay-per-subscription	-0.0116	-0.0177
	(0.0584)	(0.0346)
Constant	0.1312***	0.6511***
	(0.0121)	(0.0073)
R^2 (within)	0.008	0.001
Observations	3,216	3,216
Individuals	402	402

Notes: Fixed-effects regressions with robust standard errors clustered at the journalist level (402 clusters). The reference category is the flat-rate payment group. A joint Wald test fails to reject equality of the competition effects across incentive schemes ($p = 0.86$). *** p<0.01, ** p<0.05, * p<0.1.

Table A22: ATE on Factual Mistakes in Self-Written Headlines

	Average classification		Coder 1		Coder 2	
	(1)	(2)	(3)	(4)	(5)	(6)
Pay-per-click	0.1146***	0.1054***	0.0984**	0.0848*	0.1308***	0.1260***
	(0.0332)	(0.0315)	(0.0488)	(0.0448)	(0.0351)	(0.0337)
Pay-per-abo	0.0907***	0.0899***	0.0756***	0.0804***	0.1058***	0.0995***
	(0.0274)	(0.0277)	(0.0273)	(0.0297)	(0.0337)	(0.0336)
age		0.0013		0.0012		0.0013
		(0.0010)		(0.0010)		(0.0013)
male		0.0290		0.0065		0.0515*
		(0.0284)		(0.0392)		(0.0304)
editor		0.0074		0.0099		0.0049
		(0.0254)		(0.0291)		(0.0303)
Region (ref.: other)						
East Germany		-0.0350		-0.0299		-0.0401
		(0.0705)		(0.0693)		(0.0882)
West Germany		-0.0064		-0.0083		-0.0045
		(0.0656)		(0.0578)		(0.0827)
No answer		-0.0179		-0.0328		-0.0030
		(0.0866)		(0.0747)		(0.1169)
Political orientation (ref.: AfD)						
CDU/CSU		-0.0442		0.0167		-0.1052
		(0.1466)		(0.1142)		(0.1892)
The Greens		-0.1057		-0.0246		-0.1869
		(0.1421)		(0.1112)		(0.1832)
The Left		-0.1487		-0.0422		-0.2552
		(0.1429)		(0.1118)		(0.1844)
FDP		0.0166		0.0889		-0.0556
		(0.1602)		(0.1262)		(0.2060)
No answer		-0.0862		0.0282		-0.2006
		(0.1434)		(0.1155)		(0.1833)
SPD		-0.0803		0.0955		-0.2561
		(0.1585)		(0.1734)		(0.1863)
Other		-0.0915		-0.0070		-0.1760
		(0.1438)		(0.1137)		(0.1855)
Would not vote		0.0092		0.0540		-0.0356
		(0.1681)		(0.1346)		(0.2208)
Newsroom type (ref.: online)						
No answer		0.1122		0.0596		0.1648
		(0.1256)		(0.1361)		(0.1587)
Online: newspaper/magazine		-0.0287		-0.0512		-0.0062
		(0.0453)		(0.0485)		(0.0513)
Print: newspaper/magazine		0.0034		-0.0215		0.0282
		(0.0477)		(0.0515)		(0.0540)
Private broadcast		-0.0056		-0.0191		0.0079
		(0.0591)		(0.0637)		(0.0675)
Other		0.0113		-0.0868		0.1094
		(0.0705)		(0.0720)		(0.0921)
Public broadcast		-0.0089		0.0141		-0.0318
		(0.0628)		(0.0993)		(0.0523)
Education (ref.: other)						
Lower secondary (Hauptschule)		-0.0154		0.0167		-0.0476
		(0.1671)		(0.1130)		(0.2292)
No answer		-0.1121*		-0.0871		-0.1370
		(0.0660)		(0.0539)		(0.1086)
Intermediate (Realschule)		-0.0379		-0.0126		-0.0632
		(0.0768)		(0.0803)		(0.0925)
University/FH		-0.0376		0.0009		-0.0761**
		(0.0295)		(0.0339)		(0.0365)
Topic Fixed Effects	no	yes	no	yes	no	yes
Constant	0.0839***	0.1107	0.0581***	-0.0305	0.1097***	0.2519
	(0.0143)	(0.1567)	(0.0144)	(0.1302)	(0.0184)	(0.2031)
R^2	0.0210	0.0761	0.0080	0.0356	0.0231	0.1083
Observations	804	804	804	804	804	804

Notes: OLS estimates of the average treatment effects on factual mistakes in self-written headlines. Columns (1)-(2) use the average classification of two independent human coders. Columns (3)-(4) and (5)-(6) report results for each coder separately. Robust standard errors (in parentheses) are clustered at the journalist level (402 clusters). The flat-rate (fixed pay) group is the reference category. Columns (2), (4), and (6) include controls for age, male, region, political orientation (ref.: AfD), newsroom type, editor status, education, and *topic fixed effects*. *** p<0.01, ** p<0.05, * p<0.1.

Table A23: ATE on Sentiment and Emotionality by Topic

	Sentiment				Emotionality				N	Clusters
	PPC	(SE)	PPA	(SE)	PPC	(SE)	PPA	(SE)		
Politics	-0.296***	(0.101)	-0.180**	(0.091)	0.132**	(0.059)	0.087	(0.054)	665	133
Science	-0.261***	(0.087)	-0.326***	(0.101)	0.120***	(0.044)	0.124***	(0.042)	705	141
Economics	0.045	(0.102)	-0.041	(0.108)	0.087*	(0.051)	0.029	(0.058)	670	134
Technology	-0.148*	(0.085)	-0.110	(0.089)	0.163**	(0.065)	0.076	(0.068)	645	129
Consumption	-0.143	(0.094)	-0.235**	(0.097)	0.092*	(0.050)	0.107**	(0.054)	655	131
Crime	-0.167	(0.115)	-0.314***	(0.115)	0.102**	(0.047)	0.039	(0.054)	680	136

Joint Wald tests for equality across topics:
Pay-per-click: $F(5, 401) = 0.33$, $p = 0.89$ Pay-per-subscription: $F(5, 401) = 0.17$, $p = 0.98$

Notes: Each row reports coefficients from separate OLS regressions of sentiment and emotionality on indicators for pay-per-click (PPC) and pay-per-subscription (PPA) incentives, relative to flat pay. Standard errors (in parentheses) are robust and clustered at the journalist level. Each journalist contributed headline decisions for two of six articles. Consequently, the number of observations refers to the total number of article–headline decisions, and the number of clusters corresponds to the number of unique journalists in each regression.
Joint Wald tests refer to pooled models with topic-by-treatment interactions. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A24: Bonferroni Correction for Primary Outcomes

	Emotionality		Sentiment	
	Raw p-value	Bonferroni p-value	Raw p-value	Bonferroni p-value
Pay-per-click	0.00001271	0.00002542	0.00029992	0.00059983
Pay-per-abo	0.00604664	0.01209328	0.00004964	0.00009928

Notes: This table reports raw p-values and Bonferroni-adjusted p-values for the two pre-registered primary outcomes (sentiment and emotionality). P-values correspond to the treatment effects of the Pay-per-click and Pay-per-subscription incentive schemes from regressions with robust standard errors clustered at the journalist level (402 clusters). The Bonferroni adjustment multiplies each raw p-value by two, reflecting family-wise error control across the two primary outcomes. All treatment effects remain statistically significant after correction.

Table A25: Robustness: Treatment Effects Using Only the First (Non-Competitive) Headline Decision

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	-0.1564***	-0.1685***	0.1532***	0.1648***
	(0.0560)	(0.0520)	(0.0320)	(0.0314)
Pay-per-abo	-0.2015***	-0.2198***	0.0880**	0.0827**
	(0.0583)	(0.0551)	(0.0353)	(0.0339)
age		0.0022		-0.0011
		(0.0020)		(0.0011)
male		-0.0861*		-0.0011
		(0.0457)		(0.0271)
editor		0.0136		-0.0110
		(0.0465)		(0.0282)
Region (ref.: other)				
East Germany		-0.1593		-0.0644
		(0.1303)		(0.0890)
West Germany		-0.1559		0.0052
		(0.1205)		(0.0820)
No answer		-0.3798**		-0.0426
		(0.1752)		(0.1365)
Political orientation (ref.: AfD)				
CDU/CSU		0.5435***		-0.2602**
		(0.1534)		(0.1208)
The Greens		0.6331***		-0.2064*
		(0.1360)		(0.1128)
The Left		0.6415***		-0.2109*
		(0.1454)		(0.1197)
FDP		0.4697**		-0.1310
		(0.1904)		(0.1241)
SPD		0.4524***		-0.2289**
		(0.1417)		(0.1157)
Other		0.5052***		-0.1976
		(0.1518)		(0.1199)
Would not vote		0.5253***		-0.2220*
		(0.1420)		(0.1171)
Newsroom type (ref.: online)				
No answer		-0.0055		0.0905
		(0.1236)		(0.0751)
Print		-0.0945		0.0439
		(0.0800)		(0.0509)
Private broadcast		-0.0398		0.0463
		(0.1218)		(0.0621)
Public broadcast		-0.0240		-0.0234
		(0.0817)		(0.0526)
Other		-0.0909		0.0197
		(0.1152)		(0.0845)
Education (ref.: other)				
Lower secondary (Hauptschule)		-0.4817		-0.0881
		(0.3207)		(0.1440)
No answer		0.2592		-0.2604***
		(0.2044)		(0.0917)
Intermediate (Realschule)		0.3586***		-0.1639
		(0.1353)		(0.1073)
University		0.0757		-0.0586*
		(0.0587)		(0.0306)
Topic fixed effects	no	yes	no	yes
Constant	0.2403***	-0.2186	0.5758***	0.9213***
	(0.0337)	(0.1888)	(0.0229)	(0.1428)
R^2	0.0123	0.1155	0.0185	0.1203
Observations	1,608	1,608	1,608	1,608

Notes: This table reports robustness checks using only the first headline decision (non-competitive context), eliminating any possibility of spillovers or order effects from repeated tasks. Estimates are OLS with robust standard errors clustered at the journalist level (402 clusters). The flat-pay condition is the reference category. *** p<0.01, ** p<0.05, * p<0.1.

Table A26: Robustness: Two-Way Clustering by Journalist and Topic

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	-0.1623**	-0.1706**	0.1109***	0.1220***
	(0.0478)	(0.0448)	(0.0137)	(0.0153)
Pay-per-abo	-0.1980**	-0.2092***	0.0737**	0.0737***
	(0.0491)	(0.0393)	(0.0185)	(0.0163)
age		0.0008		0.0002
		(0.0028)		(0.0009)
male		-0.0814		-0.0120
		(0.0540)		(0.0248)
editor		-0.0330		-0.0083
		(0.0302)		(0.0129)
Region (ref.: other)				
East Germany		-0.1699		-0.0452
		(0.1077)		(0.0458)
West Germany		-0.2071		0.0144
		(0.1144)		(0.0535)
No answer		-0.2788		-0.0212
		(0.2033)		(0.0983)
Political orientation (ref.: AfD)				
CDU/CSU		0.4958***		-0.2289*
		(0.0871)		(0.0969)
The Greens		0.5731**		-0.2041
		(0.1465)		(0.0992)
The Left		0.5185**		-0.2342
		(0.1214)		(0.1154)
FDP		0.4428**		-0.2128
		(0.1677)		(0.0950)
No answer		0.3152		-0.1330
		(0.1361)		(0.1622)
SPD		0.4534**		-0.2470**
		(0.1168)		(0.0784)
Other		0.5324**		-0.1872
		(0.1583)		(0.1006)
Would not vote		0.5719**		-0.2434*
		(0.1595)		(0.0938)
Newsroom type (ref.: online)				
No answer		-0.1291		0.1267**
		(0.1582)		(0.0461)
Print		-0.1067		0.0458
		(0.0996)		(0.0439)
Private broadcast		-0.0569		0.0695
		(0.1102)		(0.0379)
Public broadcast		-0.0816		0.0057
		(0.0474)		(0.0586)
Other		-0.0650		-0.0179
		(0.1051)		(0.0799)
Education (ref.: other)				
Lower secondary (Hauptschule)		-0.3877		0.0322
		(0.3166)		(0.1458)
No answer		0.2181		-0.1754**
		(0.1540)		(0.0615)
Intermediate (Realschule)		0.4045**		-0.1365
		(0.1252)		(0.0844)
University		0.0958		-0.0469*
		(0.0480)		(0.0202)
Topic fixed effects	no	yes	no	yes
Constant	0.1265	-0.2429	0.6013***	0.9201***
	(0.0742)	(0.0851)	(0.0578)	(0.1344)
R^2	0.0118	0.0719	0.0101	0.0947
Observations	4,020	4,020	4,020	4,020

Notes: This table reports OLS estimates with two-way clustered standard errors at the levels of journalist (402 clusters) and article topic (6 clusters). The flat-pay group is the reference category. Columns (2) and (4) include the full set of demographic, political, education, newsroom-type, and topic controls. *** p<0.01, ** p<0.05, * p<0.1.

Table A27: Robustness: Reclassification of Free-Text Headlines Using RoBERTa

	Sentiment		Emotionality	
	(1)	(2)	(3)	(4)
Pay-per-click	0.0021	-0.0107	0.0062	0.0237
	(0.0400)	(0.0405)	(0.0363)	(0.0354)
Pay-per-abo	0.0103	0.0064	0.0135	0.0241
	(0.0428)	(0.0425)	(0.0390)	(0.0373)
age		0.0028*		-0.0026**
		(0.0016)		(0.0013)
male		-0.0290		0.0211
		(0.0371)		(0.0320)
editor		-0.0101		0.0194
		(0.0381)		(0.0315)
Region (ref.: other)				
East Germany		-0.1036		0.0647
		(0.1293)		(0.0879)
West Germany		-0.0458		-0.0649
		(0.1106)		(0.0763)
No answer		0.0345		-0.1667
		(0.1443)		(0.1168)
Political orientation (ref.: AfD)				
CDU/CSU		0.1815		-0.0907
		(0.2151)		(0.1946)
The Greens		0.2319		-0.1460
		(0.2100)		(0.1927)
The Left		0.1745		-0.1210
		(0.2166)		(0.1966)
FDP		-0.0060		0.0288
		(0.2235)		(0.2066)
No answer		0.0830		-0.0624
		(0.2105)		(0.1943)
SPD		0.1970		-0.1596
		(0.2137)		(0.1971)
Other		0.1384		-0.0768
		(0.2151)		(0.1967)
Would not vote		0.2286		-0.2371
		(0.2290)		(0.2159)
Newsroom type (ref.: online)				
No answer		0.0459		-0.1312
		(0.1221)		(0.1032)
Print		-0.0023		-0.0348
		(0.0723)		(0.0549)
Private broadcast		0.1364		-0.1009
		(0.0880)		(0.0694)
Public broadcast		0.0682		-0.0827
		(0.0822)		(0.0624)
Other		0.0529		-0.0712
		(0.0970)		(0.0824)
Education (ref.: other)				
Lower secondary		-0.1158		0.0272
		(0.1576)		(0.1218)
No answer		-0.1277		0.1111
		(0.1190)		(0.1185)
Intermediate		0.0581		-0.1025
		(0.0973)		(0.0902)
University		-0.0465		0.0432
		(0.0422)		(0.0324)
Topic fixed effects	no	yes	no	yes
Constant	-0.2387***	-0.5207**	0.2839***	0.5784**
	(0.0256)	(0.2320)	(0.0242)	(0.2026)
R^2	0.0001	0.0734	0.0001	0.1492
Observations	804	804	804	804

Notes: This table reports OLS estimates for the subset of self-written (free-text) headlines reclassified using a RoBERTa sentiment model. Standard errors are clustered at the journalist level (402 clusters). The flat-pay condition is the reference category. None of the treatment effects on sentiment or emotionality are statistically significant under this alternative classification.

A.2.2 Journalists' experiment 1 (Pilot)

A first online experiment with N=201 professional journalists was conducted between November 17 and December 14, 2021, in cooperation with a German journalist association. Participation was restricted to full-time journalists. The survey was distributed via email to association members, who could participate at any time during the data collection period. The experiment was implemented using the survey software *Qualtrics*, and the median completion time was 6.3 minutes. Participants were compensated either based on the click rates of the headlines they selected or with a flat rate. The average payment was €7.73, and payments were processed via PayPal.

In addition to the experiment with journalists, a survey experiment with a sample of N=299 "readers" was conducted. Further details on this connected experiment are provided in Appendix A.3.2. Ethical approval for both experiments was granted by the Ethics Committee of the Faculty of Management, Economics and Social Sciences at the University of Cologne (reference: 210036LM). The studies were pre-registered in the AEA Social Science Registry under AEARCTR-0008658.[10]

A.2.2.1 Setting

Journalists were asked to select a headline for a real article reporting on an economic forecast for the German economy. To interpret their decisions and readers' reactions, it is important to consider the economic context in Germany at that time. In November and December 2021, the country was experiencing its fourth COVID-19 wave, while the national vaccination campaign was stagnating. The German stock index (DAX) had just reached a new all-time high above 16,000 points. Analysts expected the pandemic and potential new lockdowns to dampen economic activity, but not to cause a severe downturn. Media coverage often emphasized that the sectors most affected by lockdowns were not central to Germany's GDP growth.[11] Most forecasts predicted continued, albeit modest, economic growth.[12]

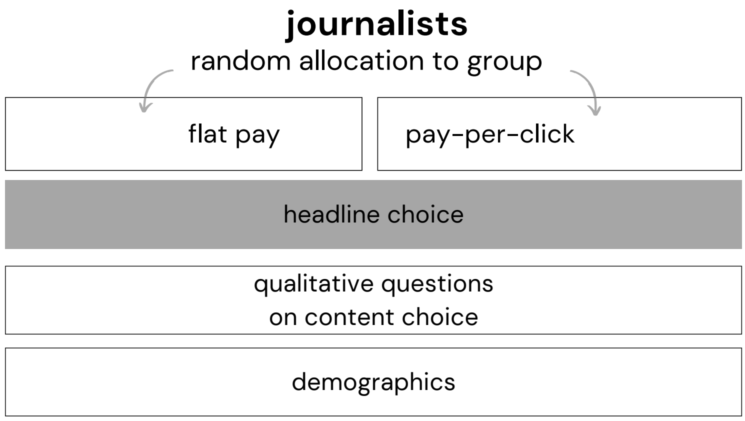
A.2.2.2 Experimental Design

The experimental setup consists of two separate experiments: one with journalists and one with readers. The two experiments are independent of each other, except for the payment scheme of the journalist treatment group, whose compensation was determined by the click rates generated by the readers. The procedures of the

[10]The pre-registration is available here.
[11]For example, see a report from *Tagesschau*.
[12]See, for instance, reports from *Die Zeit* and *Süddeutsche Zeitung*.

journalist experiment are described in the following paragraphs, while the reader experiment is detailed in Section A.3.2. Figure A17 provides an overview of the design of the journalist experiment.

Figure A17: Experimental Procedures – Overview



Notes: Figure A17 gives an overview of the experimental procedures. Randomization is indicated with arrows. The main outcome is shaded in grey.

In the journalist experiment, participants were randomly assigned by the computer to one of two groups: the “flat pay” group or the “pay-per-click” group. Each journalist was then asked to select one headline out of three options for a given article. The article was a real report by the news agency *dpa* on the expected future development of the German economy.[13] The three headline options were all factually accurate but differed in their sentiment: one was positive, one neutral, and one negative.[14] The complete wording of the article and the headline options, as well as their English translation, are provided on the next pages (Figures A18 and A19).

The “flat-pay” group received a fixed payment for completing the task, whereas the remuneration of the treatment group depended on how often their selected headline was clicked on by readers. After selecting a headline, journalists had the option to propose an alternative headline in a free-text field.[15] Participants were also asked to indicate which factors they considered relevant when choosing a headline for an article (qualitative questions”). Finally, demographic and preference information was collected, including seniority, age, political orientation, and economic preferences (using a short form of the survey modules proposed by Falk et al., 2018). A translation

[13]An article from a news agency was chosen because such reports are typically concise and relatively neutral. They are not published directly by the agency but provided as a service to newsrooms, which often adapt them slightly before publication. Versions of the article used in the experiment appeared, for example, in *Tagesschau*, *Zeit Online*, and *Focus Online*, among others.

[14]All algorithms discussed in Appendix A.1.1 classify the positive headline as positive, the neutral one as neutral, and the negative one as negative.

[15]This part of the experiment was not incentivized.

of the full experimental instructions is provided in Appendix A.4.1.

Figure A18: Given article in experiment – German

"In ihrem im Oktober veröffentlichten Herbstgutachten gehen führende Wirtschaftsforschungsinstitute von einem Wachstum des Bruttoinlandsprodukts um 2,4 Prozent in 2021 aus. Im Frühjahr hatten sie noch damit gerechnet, dass in diesem Jahr ein Anstieg um 3,7 Prozent zu erwarten sei.

Die wirtschaftliche Lage in Deutschland sei nach wie vor von der Coronapandemie gekennzeichnet, hieß es. Im Verlauf des Jahres 2022 dürfte die deutsche Wirtschaft aber wieder die Normalauslastung erreichen. Laut Prognose der Institute steigt das Bruttoinlandsprodukt im Jahr 2022 um 4,8 Prozent. In ihrer Frühjahrsprognose gingen die Institute nur von einem Plus um 3,9 Prozent für das nächste Jahr aus."

Quelle: dpa

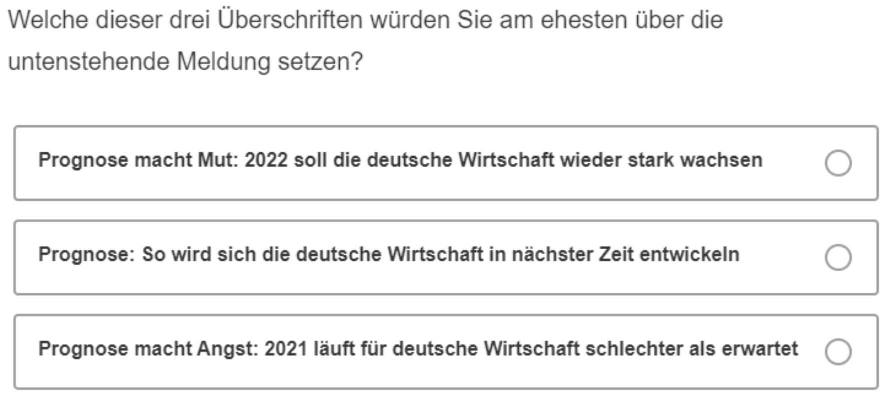
Notes: Figure A18 contains the text of the article as it was displayed to journalists and readers. A English translation can be found below. This article has originally been a short report of the German news agency *dpa*.

English translation

In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.

The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.

Figure A19: Possible headlines for given article – German


Welche dieser drei Überschriften würden Sie am ehesten über die untenstehende Meldung setzen?

Prognose macht Mut: 2022 soll die deutsche Wirtschaft wieder stark wachsen

Prognose: So wird sich die deutsche Wirtschaft in nächster Zeit entwickeln

Prognose macht Angst: 2021 läuft für deutsche Wirtschaft schlechter als erwartet

Notes: Figure A19 contains the original wording of the available headline choices. A translation to English can be found below. The headlines have been compiled by the researcher and evaluated by 10 professional journalists as being in principle realistic formulations for a headline.

English translation

Which of the following headlines would you most likely put above the article below?

- Encouraging forecast: the German economy is expected to grow strongly again in 2022
- Forecast: This is how the German economy will develop in the near future
- Scary forecast: 2021 will be worse than expected for the German economy

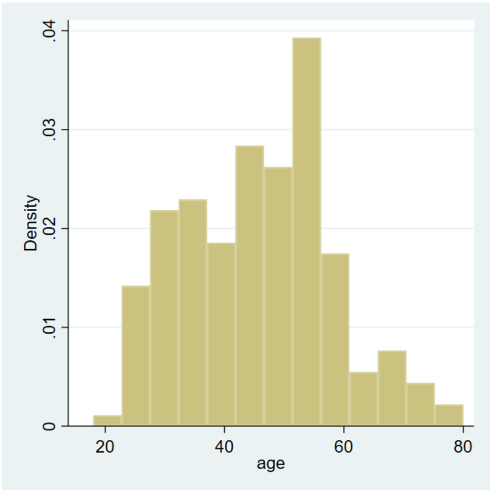
Complete experimental instructions are available in Appendix A.4.1.

A.2.2.3 Sample Descriptives

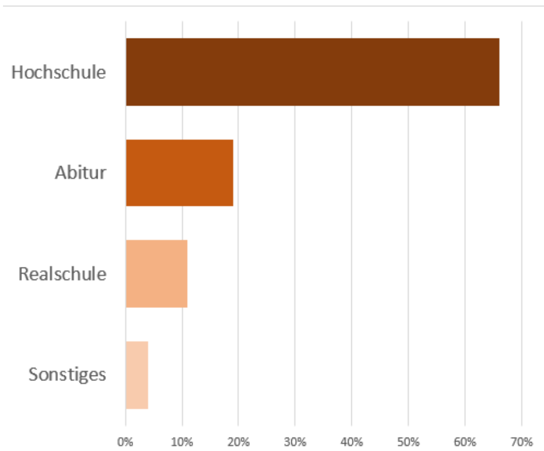
The journalist sample consists of 201 participants aged between 18 and 80 years, with between two and 42 years of professional experience in the news industry. Compared to the overall German population, political preferences in the sample are skewed toward left-leaning parties, consistent with previous survey evidence among journalists. The distribution of covariates, as well as a comparison of political preferences with those of the general population, is provided below in Figures A20 and A21.

Figure A20: Age and Education Distributions of Journalists

(a) age

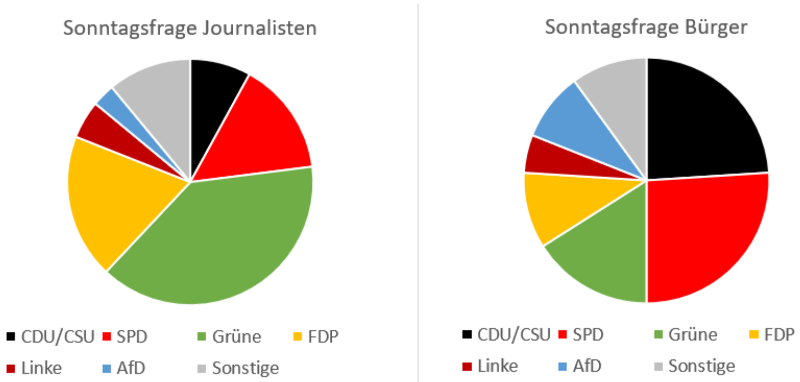


(b) education



Notes: Figure A20 illustrates the age distribution (left) and the education levels (right) of the journalist sample.

Figure A21: Voting intentions



Notes: Figure A21 summarizes answers to the question of which party one would vote for if there was an election on the following Sunday for the journalist sample (left) and a representative sample for the German population (right). The data of the representative sample comes from the market research institute *forsa*.

The randomization of journalists into the treatment and control groups was successful. Out of the 20 covariates considered, only one differs statistically significantly between the two groups, as would be expected by chance. Specifically, participants in the flat-pay group reported significantly more often that they vote for parties not currently represented in the German parliament. Controlling for this imbalance does not affect the results. The balance table for this sample is presented in Table A28.

Table A28: Balance Table Journalists

	flat pay (1)	pay-per-click (2)	t-test (3)
age	46.830	46.179	0.651
	(1.226)	(1.317)	
seniority	15.217	14.284	0.933
	(0.914)	(0.964)	
risk	7.613	7.968	-0.355
	(0.190)	(0.212)	
patience	8.151	8.179	-0.028
	(0.169)	(0.205)	
altruism	8.575	8.979	-0.403
	(0.186)	(0.189)	
trust	5.226	4.916	0.311
	(0.216)	(0.247)	
narcissism	2.679	2.937	-0.258
	(0.123)	(0.152)	
education			
secondary school	9.4%	12.6%	-3.2%
high school	23.6%	16.8%	6.7%
university	64.2%	67.4%	-3.2%
other	2.8%	3.2%	-0.3%
political preference			
SPD	15.1%	12.6%	2.5%
CDU/CSU	6.6%	9.5%	-2.9%
Die Grünen	36.8%	36.8%	0%
FDP	17.0%	20.0%	-3.0%
AfD	1.9%	3.2%	-1.3%
Die Linke	3.8%	5.3%	-1.5%
other	16%	6.3%	9.7%**
wouldn't vote	2.8%	6.3%	-3.5%
phone use	31.1%	32.6%	-1.5%
online medium	33.0%	29.3%	3.7%
Observations	106	95	

Notes: seniority: aged worked as a journalist; education: highest educational diploma; political preference: percentage of people who would vote for a certain party if there were national elections on the next Sunday; risk, patience and altruism: survey measures of these preferences validated by Falk et al., 2018 on a 11-point Likert scale; narcissism: self-evaluated narcissism on a 7-point Likert scale; phone use: percentage of participants answering the survey on their phone; online medium: percentage of participants working only for an online medium.
The value displayed for t-tests in column 3 are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

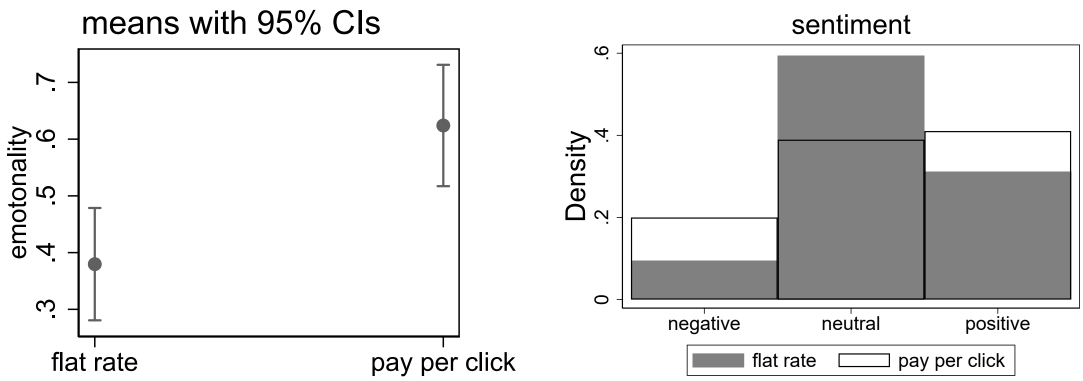
A.2.2.4 Results

Headline choice Figure A22a shows the mean headline choices of journalists in the two groups. Journalists who were paid per click were on average 20 percent more likely to choose an emotional headline than those who received a flat payment. This corresponds to a difference of 0.41 standard deviations ($p = 0.001$). As illustrated by the distribution in Figure A22b, this effect is driven by pay-per-click journalists selecting both the negative and the positive headlines more frequently than participants in the flat-pay group.

Figure A22a depicts the means of the headline choices of the journalists in the respective groups. Journalists who are paid per click are on average 20 percent more likely to choose an emotional headline than the journalists who received a flat payment (this corresponds to 0.41 standard deviations, p= 0.001). As can be seen by the distribution in Figure A22b this effect is driven by pay-per-click journalists selecting both the negative and the positive headline more often than the flat-pay group.

Figure A22: Headline Choice by Journalists

(a) emotionality of choice by group **(b)** distribution of choice



Notes: Figure A22a illustrates means in the "flat rate" and the "pay per click" group in terms of emotionality. The emotionality measure is 0 if a journalist selected the neutral headline and equals 1 if the journalist selected the positive or the negative headline. Figure A22b provides an overview of the distribution of the headline choices. The choices of the "flat rate" group are shaded in gray.

Table A29 reports the main regression results of the first journalist experiment. In columns (1) and (2), the outcome variable is a sentiment measure that equals 1 if the positive headline was chosen, 0 if the neutral headline was chosen, and -1 if the negative headline was chosen. In columns (3) and (4), the outcome is a dummy for headline emotionality that equals 1 if either the negative or the positive headline was chosen, and 0 otherwise. The reference group is always the flat-pay group. Columns (1) and (3) present univariate linear regressions, while columns (2) and (4) include the full set of covariates.

Journalists remain about 20 percent more likely to choose an emotional headline when controlling for observables, and the coefficients for this effect are statistically significant at the one-percent level. For sentiment, no statistically significant effect is detected. The point estimates are small and vary in sign depending on whether controls are included. This apparent null effect arises because the positive and negative choices in the sentiment measure tend to offset each other on average.

Table A29: OLS Estimates – ATE on Headline Choice

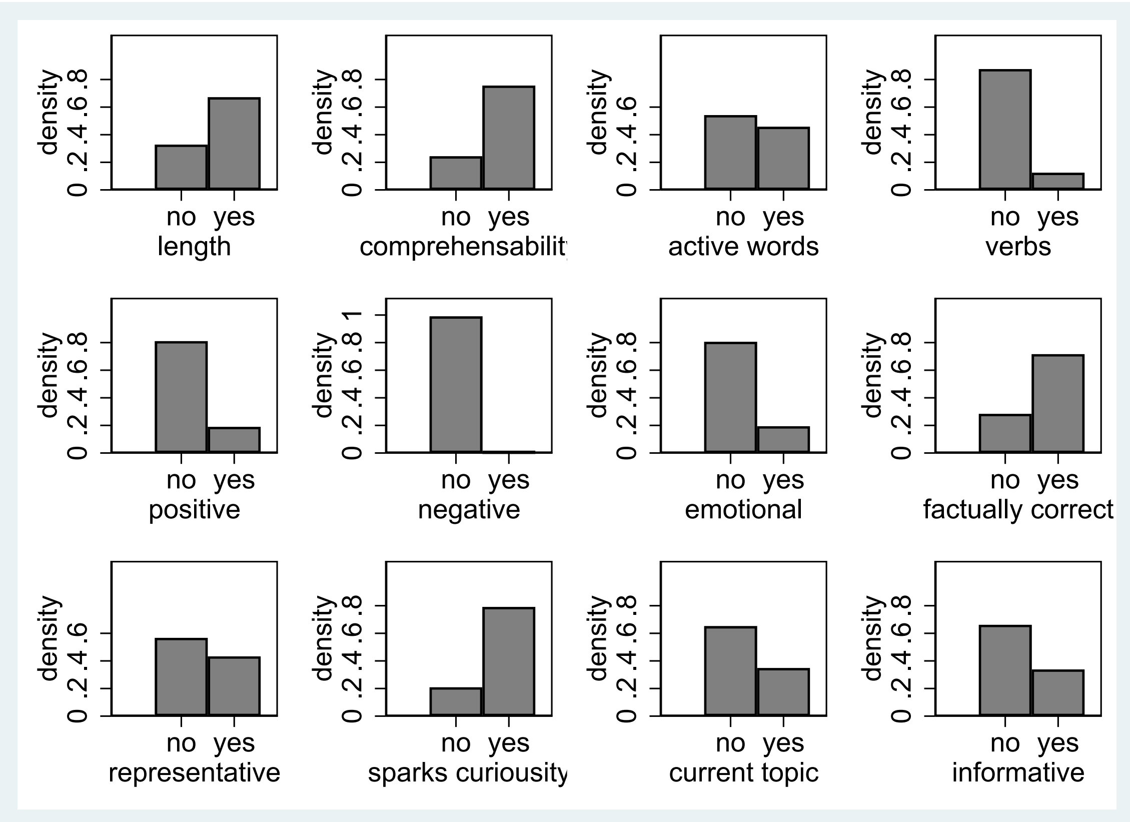
	Sentiment		Emotions	
	(1)	(2)	(3)	(4)
sentiment	-0.0064548	0.0058872		
	(0.0971431)	(0.1000453)		
emotionality			0.2048659***	0.2082093***
			(0.0694661)	(0.070733)
seniority		0.0066359		0.0055346
		(0.0063493)		(0.0047168)
age		0.0072019		0.0036173
		(0.0052546)		(0.0039627)
vote: SPD		0.0265586		0.2931497
		(0.2956812)		(0.2892631)
vote: CDU/CSU		0.1154226		0.2760218
		(0.2991308)		(0.3052485)
vote: Die Grünen		0.076108		0.444624
		(0.2815415)		(0.287009)
vote: FDP		0.0645764		0.2155939
		(0.276671)		(0.2828545)
vote: Die Linke		0.1567016		0.1004916
		(0.3168345)		(0.3139044)
vote: other		-0.1524632		0.2851998
		(0.312055)		(0.2988047)
vote: wouldn't vote		-0.160409		0.4430629
		(0.3906121)		(0.3253319)
role: producer		-0.0316488		0.0754752
		(0.1369495)		(0.1224299)
role: graphs		0.1340798		0.154554
		(0.1634986)		(0.1400021)
role: other		-0.2236172		0.0432215
		(0.2368406)		(0.1883472)
education: university		-0.0340787		0.1098455
		(0.1208012)		(0.0854872)
education: high school		0.2075383		0.3093532**
		(0.2036895)		(0.1417009)
education: other		0.1298381		0.0022261
		(0.2124546)		(0.180188)
phone use		0.069182		0.0551961
		(0.1097543)		(0.0778126)
risk preference		0.0257794		0.0297685
		(0.0287157)		(0.0198317)
patience		0.025475		0.0200179
		(0.0294635)		(0.01995)
altruism		0.0003739		-0.0131103
		(0.0294723)		(0.0231197)
trust		0.0456064*		0.014938
		(0.0231822)		(0.0165412
narcissism		-0.0629251		-0.0221183
		(0.038211)		(0.0259116)
Constant	0.2169811***	-0.7394713	0.4056604***	-0.6645239
	(0.0584536)	(0.5248686)	(0.0479311)	(0.4230708)
R^2	0.0000	0.1280	0.0418	0.1555
Observations	201	201	201	201

Notes: Table A29 reports OLS estimates with robust standard errors in parentheses. The flat rate group is the reference group. For vote the omitted category is voting for the AfD. For role the omitted category is editing journalistic pieces. For education the omitted category is A-levels (Abitur). *** p<0.01, ** p<0.05, * p<0.1

Free-text headline The text data from the free-text headline responses were analyzed using the sentiment algorithms described in Section 2.2 and through manual coding by a research assistant. Neither method reveals statistically significant differences between the two groups. This may be due to two reasons. First, this part of the experiment was not monetarily incentivized, so differences in responses would only be expected if there were "spillover" effects from the incentivized task or if participants misunderstood the payment scheme. Second, a free-text field allows for substantial variation. Assuming variation comparable to the headlines analyzed in Section 2.2, approximately 658 journalists would have been required to detect a statistically significant effect at the five-percent level. (This number represents the minimum required sample size to detect an effect of the magnitude of the coefficient when controlling for all covariates in equation A.4. As the experiment fixed not only topic, sentiment, and length but also the article itself, this estimate is likely to be a lower bound.)

Decision factors In the qualitative question, journalists indicated that the most important factors for deciding on a headline are comprehensibility, factual accuracy, appropriate length, and the ability to spark curiosity. There is no statistically significant difference between the flat-pay and pay-per-click groups in terms of which factors journalists consider relevant for headline selection. An overview of all responses regarding these potential factors is provided in Figure A23.

Figure A23: General characteristics for headlines



Notes: Figure A23 summarizes answers to the question of which factors journalist regard most important for composing a headline *in general* (not in the context of this study or certain incentives). The distributions are across both treatment groups. There is no statistically significant difference between the two groups for any of the factors on the 5 or 1 percent level.

A.3 Supplementary Material Causal Evidence on Readers

A.3.1 Main Readers' experiment

A.3.1.1 Instrument Relevance

This appendix section documents the relevance of the instruments used in the instrumental variables (IV) analyses for knowledge and reading duration. The endogenous variable is the click indicator at the headline level. Instruments are the exogenous treatment assignments that vary emotionality and competition. The IV estimates should be interpreted as local effects among readers whose clicking behavior is induced by assignment (LATE). First stage regressions are estimated on the relevant pairwise subsamples, with standard errors clustered by participant.

Across four of the five comparisons, treatment assignment significantly raises the probability of clicking. The only exception is the comparison of emotionality without competition to neutrality without competition, where the first stage is negative. This pattern is consistent with the idea that, in the absence of competitive context, emotional headlines may trigger more skepticism or lower perceived informativeness, which can reduce clicks. Importantly, all instruments are exogenous by design, and the first stages are statistically strong in most contrasts, supporting the use of IV to recover local effects among compliers.

Table A30: First Stage Regressions: Effect of Assignment on Clicking

Comparison (subsample)	Coef.	SE	t	p	$F(1, df)$	R^2	N	Clusters
EMO/C vs. NEU/NC	0.201	0.025	7.96	0.000	63.36	0.0436	4,824	804
EMO/C vs. NEU/C	0.063	0.025	2.48	0.014	6.13	0.0040	4,878	813
EMO/NC vs. NEU/NC	−0.065	0.024	−2.76	0.006	7.61	0.0061	4,824	804
EMO/C vs. EMO/NC	0.266	0.024	11.19	0.000	125.13	0.0802	4,836	806
NEU/C vs. NEU/NC	0.138	0.025	5.45	0.000	29.71	0.0215	4,866	811

Notes: Each row reports a separate first stage regression of the click indicator on the respective treatment assignment, estimated on the pairwise subsample indicated in the first column. Coefficients are the effects of assignment on the probability of clicking. Standard errors are clustered by participant.

A.3.1.2 Balance Table

Table A31: Balance: Descriptive statistics by experimental condition

	EMO/C (1)	EMO/NC (2)	NEU/C (3)	NEU/NC (4)
Prior knowledge (Consumption)	3.04 (1.05)	3.10 (1.07)	2.94 (1.07)	3.00 (1.08)
Prior knowledge (Crime)	3.14 (0.96)	3.25 (0.84)	3.21 (0.92)	3.17 (0.91)
Prior knowledge (Economy)	3.23 (0.95)	3.31 (0.91)	3.22 (1.00)	3.18 (0.97)
Prior knowledge (Science)	2.42 (1.01)	2.42 (1.04)	2.37 (0.98)	2.34 (1.07)
Prior knowledge (Technology)	3.01 (0.94)	3.01 (0.99)	2.93 (0.93)	2.88 (1.01)
Prior knowledge (Politics)	3.15 (0.93)	3.23 (0.91)	3.10 (0.88)	3.03 (0.96)
Prior mood	67.26 (22.69)	70.05 (21.40)	68.96 (21.16)	67.29 (21.20)
Prior Opinion (Consumption)	1.34 (0.91)	1.42 (0.86)	1.40 (0.89)	1.39 (0.93)
Prior Opinion (Crime)	0.75 (1.03)	0.77 (0.96)	0.84 (0.98)	0.82 (1.00)
Prior Opinion (Economy)	0.25 (0.94)	0.25 (0.92)	0.15 (0.92)	0.09 (0.90)
Prior Opinion (Science)	0.19 (1.24)	0.14 (1.28)	0.22 (1.20)	0.36 (1.24)
Prior Opinion (Technology)	0.68 (1.01)	0.81 (1.04)	0.73 (1.10)	0.79 (1.02)
Prior Opinion (Politics)	0.53 (1.05)	0.60 (1.03)	0.50 (1.04)	0.51 (1.02)
Male	50.9%	51.9%	48.0%	47.6%
Age				
18–29 years (2)	21.3%	17.6%	16.8%	20.4%
30–39 years (3)	17.6%	18.9%	16.3%	20.2%
40–49 years (4)	17.9%	16.1%	16.8%	15.7%
50–59 years (5)	19.1%	21.3%	22.9%	23.4%
60–74 years (6)	24.1%	26.1%	27.1%	20.2%
Digital media use				
Several times per month	6.5%	7.4%	8.0%	7.7%
Several times per week	14.9%	19.6%	20.2%	18.7%
Daily, one to two hours	17.1%	18.9%	18.5%	20.4%
Daily, less than one hour	41.4%	36.0%	36.3%	31.9%
No answer	20.1%	18.1%	16.8%	21.2%
Print media use				
Never	23.8%	20.8%	21.0%	22.4%
Several times per week	14.6%	19.9%	15.9%	13.5%
Daily, less than one hour	18.6%	17.4%	21.5%	15.0%
Less than once per month	15.1%	16.4%	13.7%	20.4%
No answer	27.8%	25.6%	28.0%	28.7%
Income				
€1,500–€2,500	21.1%	18.1%	21.2%	17.5%
€2,500–€3,500	22.6%	22.6%	20.7%	20.9%
€3,500–€4,500	15.6%	21.6%	20.0%	19.2%
€4,500–€5,500	14.9%	12.2%	12.0%	14.0%
No answer	25.8%	25.6%	26.1%	28.4%
Occupation				
IT, Mathematics, Natural Sciences	10.7%	9.7%	4.6%	8.7%
Business/Administrative	16.6%	14.9%	21.5%	18.2%
Medicine and Health	8.4%	5.5%	4.9%	6.2%
Miscellaneous	24.8%	33.3%	31.2%	29.4%
No answer	39.5%	36.7%	37.8%	37.4%
Political preference				
AfD	15.1%	16.1%	19.3%	16.2%
CDU/CSU	25.3%	27.0%	25.1%	24.7%
SPD	13.2%	16.4%	13.2%	15.5%
Other parties	14.4%	11.4%	12.9%	10.2%
No answer	32.0%	29.0%	29.5%	33.4%
Education				
Upper secondary (Abitur/Fachabitur)	27.5%	27.8%	21.7%	23.9%
Lower secondary (Hauptschule)	7.4%	9.4%	8.5%	11.2%
Intermediate secondary (Realschule)	32.5%	34.5%	38.3%	34.7%
University or Univ. of Applied Sciences	32.0%	28.3%	31.0%	29.7%
Participants	403	403	410	401

Notes: Continuous variables are mean (sd); categorical variables are percentages. Opinions measured on a –2 to 2 Likert scale.

Table A32: Balance: Pairwise p-values for differences by condition

	1–2	1–3	1–4	2–3	2–4	3–4
Kn. (Consumption)	0.407	0.175	0.596	**0.030****	0.180	0.418
Kn. (Crime)	**0.068***	0.253	0.591	0.510	0.191	0.534
Kn. (Economy)	0.211	0.898	0.449	0.178	**0.045****	0.541
Kn. (Science)	1.000	0.401	0.275	0.410	0.283	0.763
Kn. (Technology)	0.971	0.193	**0.061***	0.219	**0.073***	0.520
Kn. (Politics)	0.237	0.489	**0.070***	**0.054****	**0.003*****	0.231
Prior mood	**0.074***	0.270	0.985	0.467	**0.067***	0.262
Opinion (Consumption)	0.189	0.382	0.498	0.662	0.550	0.863
Opinion (Crime)	0.804	0.244	0.393	0.339	0.525	0.761
Opinion (Economy)	0.970	0.137	**0.015****	0.142	**0.015****	0.335
Opinion (Science)	0.595	0.697	**0.045****	0.355	**0.012****	**0.097***
Opinion (Technology)	**0.092***	0.550	0.150	0.305	0.800	0.432
Opinion (Politics)	0.398	0.648	0.761	0.187	0.244	0.877
Male	0.778	0.421	0.359	0.277	0.230	0.905
Age						
18–29 (2)	0.182	0.102	0.756	0.766	0.307	0.186
30–39 (3)	0.648	0.628	0.350	0.346	0.632	0.155
40–49 (4)	0.512	0.696	0.414	0.788	0.871	0.666
50–59 (5)	0.430	0.181	0.133	0.586	0.475	0.862
60–74 (6)	0.516	0.326	0.186	0.742	**0.049****	**0.021****
Digital media use						
Monthly	0.579	0.380	0.480	0.747	0.878	0.867
Weekly	**0.076***	**0.045****	0.148	0.819	0.746	0.580
Daily 1	0.521	0.598	0.227	0.906	0.570	0.492
Daily 2	0.112	0.136	**0.005*****	0.915	0.224	0.184
No answer	0.474	0.229	0.701	0.629	0.271	0.113
Print media use						
Never	0.310	0.330	0.643	0.963	0.582	0.612
Weekly	**0.050****	0.630	0.632	0.137	**0.015****	0.337
Daily 1	0.647	0.310	0.166	0.140	0.354	**0.017****
Monthly	0.629	0.548	**0.049****	0.278	0.136	**0.010*****
No answer	0.473	0.935	0.780	0.423	0.320	0.842
Income						
€1,500–€2,500	0.287	0.964	0.191	0.265	0.807	0.175
€2,500–€3,500	1.000	0.522	0.575	0.522	0.575	0.940
€3,500–€4,500	**0.030****	0.104	0.182	0.577	0.401	0.775
€4,500–€5,500	0.257	0.219	0.710	0.928	0.447	0.393
No answer	0.936	0.925	0.403	0.861	0.359	0.456
Occupation						
IT, Maths	0.641	**0.001*****	0.352	**0.005****	0.642	**0.019****
Business	0.499	**0.079***	0.555	**0.015****	0.206	0.245
Health	**0.096***	**0.042****	0.231	0.708	0.639	0.399
Other	**0.008*****	**0.042****	0.141	0.536	0.243	0.579
No answer	0.425	0.629	0.551	0.750	0.841	0.907
Pol. preference						
AfD	0.698	0.119	0.676	0.241	0.975	0.254
CDU/CSU	0.575	0.951	0.839	0.532	0.445	0.886
SPD	0.197	0.993	0.350	0.197	0.723	0.352
Other	0.207	0.543	**0.072***	0.510	0.587	0.229
No answer	0.359	0.440	0.671	0.880	0.180	0.231
Education						
Upper secondary	0.937	**0.053***	0.243	**0.044****	0.212	0.449
Lower secondary	0.311	0.566	**0.066***	0.656	0.404	0.200
Intermediate secondary	0.550	**0.085***	0.517	0.260	0.959	0.283
University	0.250	0.751	0.474	0.401	0.664	0.687

Notes: Cells report p-values from two-sample tests: t-tests for continuous variables and proportion tests for binary/categorical indicators. Stars denote significance at conventional levels (* p<0.10, ** p<0.05, *** p<0.01). Pairwise comparisons: 1=EMO/C, 2=EMO/NC, 3=NEU/C, 4=NEU/NC.

A.3.1.3 Results: Omitted Tables

Table A33: Readers: Effect on Factual Knowledge

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	0.1029***	0.1243***	0.0790**	0.1000***	0.1270***	0.1636***	-0.0241	-0.0344	0.0239	0.0389
(Robust SE)	(0.0365)	(0.0374)	(0.0349)	(0.0344)	(0.0365)	(0.0365)	(0.0340)	(0.0341)	(0.0374)	(0.0376)
Covariates										
male		-0.0500		-0.0703*		-0.0626		-0.0989***		-0.0464
		(0.0412)		(0.0362)		(0.0414)		(0.0357)		(0.0415)
prior knowledge		-0.0828***		-0.0550***		-0.0854***		-0.0522***		-0.0786***
		(0.0167)		(0.0170)		(0.0164)		(0.0164)		(0.0169)
prior opinion		-0.0081		-0.0129		-0.0112		-0.0091		-0.0143
		(0.0148)		(0.0149)		(0.0152)		(0.0155)		(0.0146)
prior feelings		0.0007		0.0008		-0.0004		0.0002		0.0000
		(0.0008)		(0.0008)		(0.0009)		(0.0008)		(0.0009)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE)	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0026	0.0339	0.0015	0.0310	0.0041	0.0380	0.0002	0.0326	0.0001	0.0344
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Each column reports an OLS regression of log(distance to truth) (standardized to have a mean of zero and standard deviation of one) on the treatment indicator for the comparison named in the header. This means that higher values indicate lower knowledge. Robust standard errors clustered at the respondent level in parentheses. Columns with controls added include the categorial covariates for age, male, income, job type, education, digital media usage, print media usage, and political-preferences, plus the continuous pre-treatment covariates male, prior knowledge, prior opinion, and prior feelings; and article topic fixed effects. $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table A34: Readers: Effect on Factual Knowledge – Robustness

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	0.0467	0.0683**	0.0673**	0.0935***	0.0353	0.0690**	0.0115	0.0067	-0.0206	0.0011
(Robust SE)	(0.0326)	(0.0326)	(0.0318)	(0.0317)	(0.0330)	(0.0320)	(0.0319)	(0.0322)	(0.0329)	(0.0327)
Covariates										
male		-0.0565		-0.0583*		-0.0426		-0.0879***		-0.0303
		(0.0353)		(0.0329)		(0.0349)		(0.0336)		(0.0344)
prior knowledge		-0.0567***		-0.0331*		-0.0552***		-0.0239		-0.0545***
		(0.0181)		(0.0171)		(0.0184)		(0.0171)		(0.0183)
prior opinion		0.0030		-0.0116		-0.0122		-0.0179		-0.0053
		(0.0139)		(0.0141)		(0.0156)		(0.0159)		(0.0139)
prior feelings		0.0005		0.0011		-0.0009		0.0003		-0.0002
		(0.0008)		(0.0007)		(0.0009)		(0.0007)		(0.0009)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0005	0.0298	0.0011	0.0253	0.0003	0.0286	0.0000	0.0224	0.0001	0.0272
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Outcome is the standardized *distance to truth* (higher values = lower factual knowledge). Each column reports an OLS regression of the outcome on the treatment indicator indicated in the header. Robust standard errors clustered at the respondent level in parentheses. Columns with controls include the full set of fixed effects (income, job, education, digital and print media use, political orientation, and article topic) and continuous pre-treatment covariates (male, prior knowledge, prior opinion, prior feelings). $^{***}\,p < 0.01$, $^{**}\,p < 0.05$, $^{*}\,p < 0.10$.

Table A35: Readers: IV Estimates for Factual Knowledge

	Main		**Emotionality**				**Competition**			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IV Estimate										
clicked	0.5130**	0.6856***	1.2593	1.7708*	-1.9413***	-1.9441***	-0.0906	-0.1263	0.1731	0.3158
(Robust SE)	(0.2147)	(0.2556)	(0.8216)	(1.0466)	(0.6222)	(0.4947)	(0.1255)	(0.1219)	(0.2851)	(0.3310)
Covariates										
male		-0.0778		-0.1031		0.0619		-0.0924***		-0.0540
		(0.0525)		(0.0680)		(0.0533)		(0.0356)		(0.0461)
prior knowledge		-0.1028***		-0.1027***		-0.0136		-0.0497***		-0.0929***
		(0.0209)		(0.0376)		(0.0252)		(0.0162)		(0.0231)
prior opinion		0.0021		0.0289		-0.0061		-0.0098		-0.0092
		(0.0171)		(0.0323)		(0.0150)		(0.0151)		(0.0165)
prior feelings		0.00035		0.00090		0.00074		0.00039		0.00023
		(0.00105)		(0.00151)		(0.00101)		(0.00073)		(0.00102)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866
Clusters	804	804	813	813	804	804	806	806	811	811
Wald χ^2	5.71	111.29	2.35	187.71	9.73	201.21	0.52	940.90	0.37	304.41
First stage F	63.36	63.36	6.13	6.13	7.61	7.61	125.13	125.13	29.71	29.71

Notes: Instrumental variables (2SLS) regressions of log(distance to truth) on the endogenous indicator *clicked*. The instrument is the randomized treatment assignment in each panel header (e.g., EMO_C in columns (1)–(2), EMO_NC in (5)–(6), NEU_C in (9)–(10)). Robust standard errors clustered at the respondent level in parentheses. Columns with controls include the fixed-effect blocks listed and the continuous pre-treatment covariates (prior knowledge, prior opinion, prior feelings) plus the binary indicator *male*. Higher values of the outcome indicate lower factual knowledge. First-stage statistics correspond to Table A30. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A36: Readers: OLS Estimates for Time in News Environment

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OLS Estimate										
Coefficient	8.9534***	8.5802***	0.8913	0.9470	-0.4200	-0.8898	9.3734***	9.8367***	8.0621***	7.4342***
(Robust SE)	(1.0347)	(1.0416)	(1.2023)	(1.1957)	(0.8575)	(0.8740)	(1.0311)	(1.0279)	(1.0572)	(1.0761)
Covariates										
male		-0.7244		-1.0123		0.3012		0.2820		-0.5045
		(1.1763)		(1.2584)		(0.9847)		(1.0887)		(1.1987)
prior knowledge		1.8554***		2.2099***		1.5666***		1.5170***		2.4755***
		(0.3603)		(0.4234)		(0.3181)		(0.3800)		(0.3761)
prior opinion		-0.0453		-0.0967		-0.0007		0.2055		-0.2125
		(0.2935)		(0.3491)		(0.2610)		(0.2872)		(0.3236)
prior feelings		-0.0218		-0.0143		0.0341*		0.0210		0.0070
		(0.0250)		(0.0287)		(0.0207)		(0.0239)		(0.0273)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0524	0.1290	0.0004	0.0928	0.0002	0.0577	0.0597	0.1321	0.0422	0.1136
Observations	4,729	4,729	4,783	4,783	4,824	4,824	4,741	4,741	4,866	4,866

Notes: OLS regressions of time in news environment (in seconds) on the treatment indicator displayed in each column header, estimated on the respective pairwise subsamples. Robust standard errors clustered at the respondent level in parentheses. Columns with controls include the listed fixed-effect blocks and the continuous pre-treatment covariates (*prior knowledge, prior opinion, prior feelings*) as well as the indicator *male.* $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A37: Readers: Winsorized at 99th percentile

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	10.9670***	10.7825***	0.7708	0.9608	-0.4396	-1.0178	11.4067***	12.1166***	10.1962***	9.4213***
(Robust SE)	(1.4185)	(1.4955)	(1.7477)	(1.7697)	(1.1168)	(1.1357)	(1.4405)	(1.4855)	(1.4922)	(1.4908)
Covariates										
male		-0.7591		-1.4858		0.3118		0.3597		-0.7377
		(1.6667)		(1.8796)		(1.2715)		(1.5613)		(1.6435)
age		1.6038***		1.9143***		0.5437		1.5317***		1.0091*
		(0.4843)		(0.6313)		(0.4032)		(0.4981)		(0.5362)
prior knowledge		2.2364***		2.5566***		1.9692***		1.8939***		2.8887***
		(0.4980)		(0.6116)		(0.4360)		(0.5545)		(0.5252)
prior opinion		-0.0569		-0.1362		-0.0631		0.2475		-0.3139
		(0.4203)		(0.5319)		(0.3808)		(0.4341)		(0.4866)
prior feelings		-0.0555		-0.0290		0.0342		0.0026		0.0082
		(0.0369)		(0.0435)		(0.0293)		(0.0344)		(0.0406)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0372	0.0974	0.0001	0.0709	0.0001	0.0411	0.0409	0.0962	0.0319	0.0870
Observations	4,729	4,729	4,783	4,783	4,824	4,824	4,741	4,741	4,866	4,866

Notes: OLS regressions of time in news environment (in seconds, winsorized at the 99th percentile) on the treatment indicator shown in the column header, estimated on the respective pairwise subsample. Robust standard errors clustered at the respondent level in parentheses. Columns with controls include the listed fixed-effect blocks and continuous pre-treatment covariates (*age, prior knowledge, prior opinion, prior feelings*) as well as the indicator *male.* *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A38: Readers: Winsorized at 90th percentile

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	7.6126***	7.2164***	0.9976	1.0579	-0.2828	-0.7063	7.8953***	8.2544***	6.6149***	6.0415***
(Robust SE)	(0.8168)	(0.8113)	(0.9119)	(0.9002)	(0.6980)	(0.7098)	(0.8076)	(0.7978)	(0.8164)	(0.8301)
Covariates										
male		-0.5620		-0.7253		0.3578		0.1774		-0.2407
		(0.9196)		(0.9518)		(0.7948)		(0.8445)		(0.9377)
age		1.3629***		1.5011***		0.7066***		1.2616***		0.9917***
		(0.2889)		(0.3291)		(0.2596)		(0.2905)		(0.2989)
prior knowledge		1.5203***		1.7998***		1.2930***		1.2035***		2.0727***
		(0.2852)		(0.3228)		(0.2567)		(0.2963)		(0.2874)
prior opinion		-0.0270		-0.0090		0.0418		0.1645		-0.0671
		(0.2273)		(0.2612)		(0.2075)		(0.2182)		(0.2486)
prior feelings		-0.0114		-0.0077		0.0282*		0.0238		0.0064
		(0.0189)		(0.0215)		(0.0162)		(0.0185)		(0.0204)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0622	0.1442	0.0010	0.1005	0.0001	0.0636	0.0696	0.1492	0.0481	0.1238
Observations	4,729	4,729	4,783	4,783	4,824	4,824	4,741	4,741	4,866	4,866

Notes: OLS regressions of time in news environment (in seconds, winsorized at the 90th percentile) on the treatment indicator shown in each header, estimated on the corresponding pairwise subsample. Robust standard errors clustered at the respondent level in parentheses. Columns with controls include the fixed-effect blocks listed and the continuous pre-treatment covariates (*age, prior knowledge, prior opinion, prior feelings*) as well as the binary indicator *male*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A39: Readers: Effect on Opinion Polarization

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	-0.0236	-0.0438	-0.0327	-0.0266	-0.0027	-0.0310	-0.0208	-0.0170	0.0092	-0.0150
(Robust SE)	(0.0303)	(0.0287)	(0.0298)	(0.0283)	(0.0299)	(0.0282)	(0.0296)	(0.0282)	(0.0301)	(0.0285)
Covariates										
male		0.0436		0.0036		0.0326		0.0373		0.0001
		(0.0306)		(0.0294)		(0.0302)		(0.0299)		(0.0309)
age		0.0561***		0.0546***		0.0593***		0.0687***		0.0489***
		(0.0102)		(0.0100)		(0.0107)		(0.0101)		(0.0104)
prior knowledge		0.0979***		0.0960***		0.0973***		0.0985***		0.0899***
		(0.0136)		(0.0140)		(0.0140)		(0.0144)		(0.0133)
prior opinion		0.0870***		0.0743***		0.0663***		0.0642***		0.0789***
		(0.0115)		(0.0116)		(0.0121)		(0.0116)		(0.0120)
prior feelings		0.0002		0.0003		0.0003		-0.0005		0.0013*
		(0.0007)		(0.0007)		(0.0007)		(0.0007)		(0.0007)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0002	0.1409	0.0005	0.1408	0.0000	0.1319	0.0002	0.1273	0.0000	0.1429
Observations	4,803	4,803	4,859	4,859	4,806	4,806	4,817	4,817	4,848	4,848

Notes: Outcome is opinion polarization. Each column reports an OLS regression on the treatment indicator for the comparison shown in the header; robust standard errors clustered at the respondent level in parentheses. "With controls" columns include the FE blocks listed plus continuous pre-treatment covariates (age, prior knowledge, prior opinion, prior feelings) and the binary indicator male. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A40: Readers: Effect on Change in Opinion (absolute difference post–pre)

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	-0.0528	-0.0795**	-0.0528	-0.0466	0.0391	0.0105	-0.0918**	-0.0966**	0.0000	-0.0485
(Robust SE)	(0.0400)	(0.0397)	(0.0397)	(0.0401)	(0.0396)	(0.0379)	(0.0404)	(0.0398)	(0.0388)	(0.0379)
Covariates										
male		0.0304		0.0168		0.0781*		0.0584		0.0237
		(0.0458)		(0.0425)		(0.0425)		(0.0425)		(0.0434)
age		0.0676***		0.0811***		0.0669***		0.1011***		0.0553***
		(0.0139)		(0.0141)		(0.0150)		(0.0153)		(0.0139)
prior knowledge		0.0473**		0.0352*		0.0336*		0.0391*		0.0292
		(0.0206)		(0.0208)		(0.0204)		(0.0207)		(0.0202)
prior opinion		0.0217		0.0296		0.0260		0.0146		0.0424**
		(0.0200)		(0.0202)		(0.0201)		(0.0201)		(0.0201)
prior feelings		-0.0006		0.0005		0.0007		-0.0003		0.0014
		(0.0009)		(0.0009)		(0.0009)		(0.0010)		(0.0009)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0005	0.1078	0.0005	0.1127	0.0003	0.1155	0.0016	0.1200	0.0000	0.1059
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Outcome is the absolute change in opinion between post and pre measures. Each column reports an OLS regression on the treatment indicator for the comparison shown in the header; robust standard errors clustered at the respondent level in parentheses. "With controls" columns include the FE blocks listed plus continuous pre-treatment covariates (age, prior knowledge, prior opinion, prior feelings) and the binary indicator male. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A41: Readers: Effect on Mood (post feeling)

	Main		Emotionality				Competition			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Positive headline	-1.8418	0.2747	-2.8117	0.8289	2.8725	0.0446	-4.8133**	-0.2078		
(Robust SE)	(1.8996)	(0.6673)	(1.8937)	(0.6554)	(1.8728)	(0.5731)	(1.8747)	(0.6740)		
Negative headline	0.9056	-0.9704	-0.0642	-0.3794	3.0698*	0.9328	-2.0658	-1.4000**		
(Robust SE)	(1.8573)	(0.6302)	(1.8513)	(0.5988)	(1.6835)	(0.6394)	(1.8319)	(0.6362)		
Neutral with competition									0.9699	-0.7794
(Robust SE)									(1.4884)	(0.4844)
Covariates										
male		-0.4189		-0.1009		-0.0413		-0.2056		-0.2181
		(0.5055)		(0.5736)		(0.5325)		(0.5472)		(0.5219)
age		0.2588		-0.1398		0.2059		0.1203		-0.0911
		(0.1772)		(0.1704)		(0.2232)		(0.2137)		(0.1948)
prior knowledge		0.5741**		0.2206		0.4071**		0.5782**		0.1188
		(0.1932)		(0.1944)		(0.1857)		(0.2131)		(0.1647)
prior opinion		-0.0586		-0.1083		-0.0769		-0.2302		0.0772
		(0.1526)		(0.1387)		(0.1371)		(0.1470)		(0.1413)
prior feelings		0.9289***		0.9209***		0.9112***		0.8928***		0.9431***
		(0.0143)		(0.0142)		(0.0185)		(0.0173)		(0.0131)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0021	0.9015	0.0031	0.8916	0.0051	0.8942	0.0086	0.8856	0.0005	0.8998
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Outcome is the mood measure (post feeling). Columns (1)–(2) compare emotional headlines under competition to neutral without competition; (3)–(4) emotional competition vs. neutral competition; (5)–(6) emotional no-competition vs. neutral no-competition; (7)–(8) emotional competition vs. emotional no-competition; (9)–(10) neutral competition vs. neutral no-competition. Where applicable, two treatment indicators distinguish *positive* and *negative* emotional headlines. Robust standard errors clustered at the respondent level in parentheses. "With controls" columns include the FE blocks listed plus continuous pre-treatment covariates (age, prior knowledge, prior opinion, prior feelings) and the binary indicator male. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A42: Readers: Effect on Clicking Probability

	Main		**Emotionality**				**Competition**			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	0.2005***	0.1814***	0.0628**	0.0565**	-0.0654***	-0.0841***	0.2659***	0.2725***	0.1378***	0.1232***
(Robust SE)	(0.0252)	(0.0258)	(0.0253)	(0.0254)	(0.0237)	(0.0239)	(0.0238)	(0.0232)	(0.0253)	(0.0253)
Covariates										
male		0.0406		0.0185		0.0641**		0.0511**		0.0243
		(0.0281)		(0.0272)		(0.0262)		(0.0249)		(0.0286)
age		-0.0165*		-0.0234**		-0.0055		-0.0186**		-0.0106
		(0.0095)		(0.0096)		(0.0090)		(0.0093)		(0.0091)
prior knowledge		0.0293***		0.0269***		0.0369***		0.0200**		0.0455***
		(0.0097)		(0.0103)		(0.0089)		(0.0095)		(0.0095)
prior opinion		-0.0149**		-0.0236***		0.0027		-0.0059		-0.0162**
		(0.0074)		(0.0078)		(0.0067)		(0.0068)		(0.0076)
prior feelings		0.0004		-0.0000		0.0006		0.0016***		-0.0006
		(0.0006)		(0.0006)		(0.0005)		(0.0006)		(0.0006)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0436	0.1037	0.0040	0.0644	0.0061	0.0767	0.0802	0.1431	0.0215	0.0963
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Outcome is a binary indicator for whether the reader clicked on the article headline. Each column reports an OLS regression on the treatment indicator for the comparison shown in the header; robust standard errors clustered at the respondent level in parentheses. "With controls" columns include the FE blocks listed plus continuous pre-treatment covariates (age, prior knowledge, prior opinion, prior feelings) and the binary indicator male (coefficients printed above). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A43: Readers: Effect on Newsletter Subscription Probability

	Main		**Emotionality**				**Competition**			
	EMO C vs NEU NC		EMO C vs NEU C		EMO NC vs NEU NC		EMO C vs EMO NC		NEU C vs NEU NC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment effect										
Coefficient	-0.0251	-0.0332**	-0.0010	0.0002	0.0134	0.0038	-0.0385**	-0.0446***	-0.0241	-0.0288**
(Robust SE)	(0.0173)	(0.0162)	(0.0164)	(0.0150)	(0.0178)	(0.0158)	(0.0174)	(0.0160)	(0.0168)	(0.0144)
Covariates										
male		0.0209		0.0107		-0.0159		0.0043		-0.0022
		(0.0162)		(0.0168)		(0.0170)		(0.0169)		(0.0164)
age		-0.0208***		-0.0138**		-0.0170***		-0.0138**		-0.0179***
		(0.0060)		(0.0059)		(0.0062)		(0.0064)		(0.0058)
prior knowledge		0.0598***		0.0657***		0.0563***		0.0595***		0.0600***
		(0.0066)		(0.0069)		(0.0069)		(0.0071)		(0.0065)
prior opinion		-0.0000		-0.0034		-0.0106**		0.0026		-0.0160***
		(0.0057)		(0.0054)		(0.0054)		(0.0051)		(0.0056)
prior feelings		0.0001		0.0004		-0.0002		0.0000		0.0001
		(0.0004)		(0.0004)		(0.0004)		(0.0004)		(0.0004)
Earn FE	no	yes	no	yes	no	yes	no	yes	no	yes
Job FE	no	yes	no	yes	no	yes	no	yes	no	yes
Education FE	no	yes	no	yes	no	yes	no	yes	no	yes
Digital media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Print media FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Article Topic FE	no	yes	no	yes	no	yes	no	yes	no	yes
R^2	0.0015	0.1577	0.0000	0.1469	0.0004	0.1577	0.0033	0.1518	0.0013	0.1623
Observations	4,824	4,824	4,878	4,878	4,824	4,824	4,836	4,836	4,866	4,866

Notes: Outcome is a binary indicator for whether the reader subscribed to the newsletter. Each column reports an OLS regression on the treatment indicator for the comparison shown in the header; robust standard errors clustered at the respondent level in parentheses. "With controls" columns include the FE blocks listed plus continuous pre-treatment covariates (age, prior knowledge, prior opinion, prior feelings) and the binary indicator male (coefficients printed above). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A44: Heterogeneity in Reader Responses to Emotional Headlines

Moderator Z	Knowledge $\beta_{\text{EMO}\times Z}$	News time $\beta_{\text{EMO}\times Z}$	Polarization $\beta_{\text{EMO}\times Z}$	Post mood interactions POS$\times Z$	NEG$\times Z$	Clicked $\beta_{\text{EMO}\times Z}$	Newsletter $\beta_{\text{EMO}\times Z}$
Male	0.007	−0.645	0.013	0.023	0.542	−0.033	−0.023
Old	0.046	3.544	−0.020	−1.690	−0.689	−0.065	0.042
Left	−0.136	4.189	−0.031	1.226	−0.559	0.024	0.016
High education	0.102	−3.712	0.019	−2.368	2.930**	−0.058	−0.001
Low knowledge	−0.165**	1.014	0.028	−0.460	−0.464	0.039	0.028
Strong opinion	−0.121	2.647**	0.106*	−1.900	−0.330	0.002	0.008
Bad pre-mood	−0.103	1.197	0.052	3.038	−1.250	−0.027	−0.036
Good pre-mood	0.092	−6.096**	−0.055	−2.445	0.388	−0.086	−0.001
High digital reader	0.034	3.350	0.093	−0.329	0.280	0.013	−0.070**
High print reader	0.135	−0.823	0.003	0.200	1.188	−0.035	−0.023

Notes: Each cell reports the interaction coefficient from a separate OLS regression of the indicated outcome on an indicator for an emotional headline (EMO), the moderator Z, and their interaction, estimated on the EMO vs. neutral non-competition sample, with robust standard errors clustered at the respondent level. “Knowledge” is the log distance to the true value. “News time” is the time spent in news environment in seconds (winsorized at the 95th percentile). “Polarization” is the absolute value of the post-opinion. “Post mood interactions” report the EMO_POS$\times Z$ and EMO_NEG$\times Z$ terms from the mood specification controlling for pre-mood. Significance levels: $^{*}p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$.

Table A45: Multiple Hypothesis Correction (EMO/C vs. NEU/NC)

Outcome	Raw p-value	Bonferroni p-value
Factual knowledge	0.0009^{***}	0.0037^{**}
News Time	$7.10 \times 10^{-16***}$	$2.84 \times 10^{-15***}$
Mood	0.5221	1.0000
Opinion polarization	0.1278	0.5113

Notes: Entries report p-values for the treatment indicator (TG) from separate regressions with the standard control set and clustering by participant. Bonferroni uses $k = 4$ family size (knowledge, time in news environment, mood, polarization). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A46: Order Effects: EMO/C × Order (z-scored), Main Comparison

	(1) knowledge	**(2) news time**	**(3) opinion polarization**	**(4) post mood**
EMO_C	0.125***	8.558***	-0.044	-0.345
	(0.037)	(1.042)	(0.029)	(0.539)
Order (z)	-0.010	-0.563**	0.012	0.013
	(0.019)	(0.254)	(0.014)	(0.011)
EMO_C × Order (z)	0.045	-2.284***	-0.032	0.044
	(0.028)	(0.428)	(0.019)	(0.040)
age	-0.017	1.551***	0.056***	0.252
	(0.013)	(0.368)	(0.010)	(0.176)
Pre-knowledge	-0.083***	1.867***	0.098***	0.569***
	(0.017)	(0.360)	(0.014)	(0.194)
Pre-opinion	-0.008	-0.057	0.087***	-0.061
	(0.015)	(0.292)	(0.012)	(0.152)
Pre-feeling_5	0.001	-0.021	0.000	0.928***
	(0.001)	(0.025)	(0.001)	(0.014)
Topic FE	Yes	Yes	Yes	Yes
Other standard controls	Yes	Yes	Yes	Yes
Observations	4,824	4,729	4,803	4,824

Notes: Standard errors in parentheses, clustered at the participant level.
Dependent variables are standardized where applicable. Order corresponds to article position (1–6).
Includes all baseline controls and topic fixed effects.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

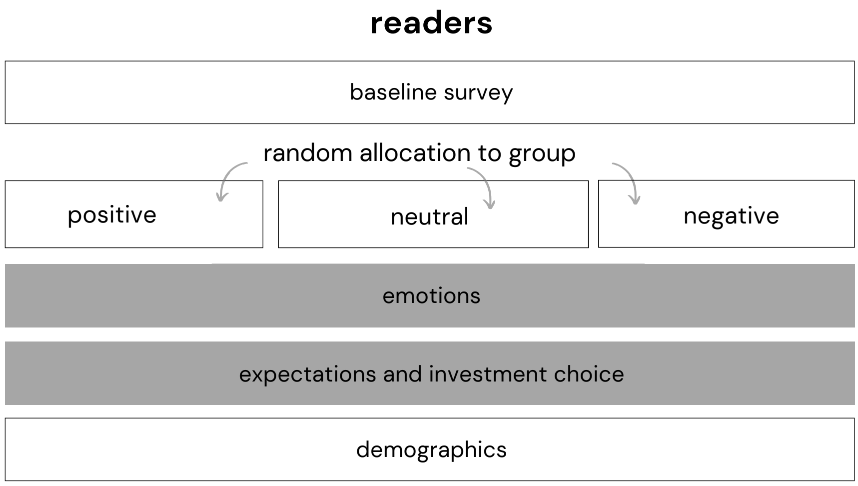
A.3.2 Readers' experiment 1 (Pilot)

The survey experiment with 299 participants was conducted during the same period as the first journalist experiment (see Appendix A.2.2). Participants were recruited from the subject pool of the Cologne Laboratory for Economic Research (CLER) using ORSEE (Greiner, 2015). The median completion time was 4.7 minutes, and participants received an average payment of € 1.98, transferred via PayPal. The experiment served two purposes: first, to generate the click-rate data used to determine payments in the journalist treatment group, and second, to provide evidence on how readers respond to different headline formulations.

A.3.2.1 Experimental Design

Readers began the experiment with a short baseline survey that elicited their current emotional state, risk preferences, and self-assessed knowledge in economics and finance. They were then randomly assigned by the computer to view one of the three headlines used in the journalist experiment.

Figure A24: Experimental Procedures – Overview



Notes: Figure A24 gives an overview of the experimental procedures. Randomization is indicated with arrows. The main outcomes are shaded in grey.

Readers could choose to click on the headline to access the full article. Each click reduced their remuneration by €0.05, corresponding to about five percent of the total payment. This small cost was introduced to discourage automatic clicking and to ensure that clicks reflected genuine interest. Participants were informed that the article content might assist them in making more informed decisions in payoff-relevant questions later in the experiment. The resulting click data were used to determine the payments for journalists in the "pay-per-click" treatment group.

After viewing the headline and deciding whether to click, participants reported their current feelings and their beliefs about the article's content, specifically regarding the expected future development of the German economy. Feelings were measured using a general question on current mood (11-point Likert scale) and the i-PANAS-sf scale (Thompson, 2007), a widely used instrument for assessing short-term affective states.[16]

Beliefs were elicited through an incentivized numerical forecasting task on expected GDP growth and the future development of the DAX index. Incentives were implemented as follows: For each variable, one participant was randomly selected and rewarded based on the accuracy of their stated forecast, with a maximum bonus of €10. Each participant could receive at most one such payoff. This design prevents hedging across forecasts and ensures a non-strategic environment, as the expected payoff does not depend on the responses of other participants. The incentive structure is therefore expected to induce truthful reporting of expectations.

Readers also participated in an incentivized investment task. They decided how much of an endowment between €0.00 and €0.50 to invest in the DAX until the end of 2021 and until the end of 2022. Any amount not invested was paid out at the same time as potential investment returns. An overview of the experimental procedures is provided in Figure A24, and a translation of the full experimental instructions for readers is included in Appendix A.4.3.

A.3.2.2 Sample Descriptives

Readers were recruited from the subject pool of the Cologne Laboratory for Economic Research and thus consist primarily of students and recent graduates. 44 percent of participants were male, and the median age was 27. The sample is broadly balanced across treatment conditions. Out of 45 balance tests, five differences are statistically significant at conventional levels, which is consistent with random variation. These differences do not follow a systematic pattern across groups, and the results remain unchanged when controlling for observable characteristics. A full balance table is provided below as Table A47.

A.3.2.3 Results

The results indicate significant responses in readers' short-term emotions and expectations, but no statistically significant effects on investment decisions or clicking

[16]The i-PANAS-sf measures affect across ten dimensions: five positive (active, inspired, determined, attentive, alert) and five negative (afraid, nervous, ashamed, hostile, upset). Each item is rated on a 5-point Likert scale, and an overall affect index is constructed by summing the positive and subtracting the negative dimensions.

Table A47: Balance Table Readers

	positive (1)	neutral (2)	negative (3)	t-test 1 VS. 2 (4)	t-test 1 VS. 3 (5)	t-test 2 VS. 3 (6)
age	27.606	26.840	28.930	0.766	-1.324	-2.090**
	(0.562)	(0.567)	(0.800)			
male	47.5%	40.0%	47.0%	7.5%	0.5%	7.0%
political preference						
SPD	7.1%	12.0%	17.0%	-4.9%	-9.9%**	-5.0%
CDU/CSU	12.1%	11.0%	7.0%	1.1%	5.1%	4.0%
Die Grünen	41.4%	36.0%	39.0%	5.4%	2.4%	-3.0%
FDP	18.2%	20.0%	16.0%	-1.8%	2.2%	4.0%
AfD	1.0%	1.0%	1.0%	0%	0%	0%
Die Linke	9.1%	4.0%	4.0%	5.1%	5.1%	0%
other	6.1%	7.0%	5.0%	0.9%	1.1%	2.0%
wouldn't vote	5.1%	9.0%	11.0%	-3.9%	-5.9%	-2.0%
phone use	7.1%	6.0%	16.0%	1.1%	-8.9%**	-10.0%**
ex-ante feeling	12.323	12.430	12.440	-0.107	-0.117	-0.010
	(0.205)	(0.212)	(0.200)			
econ knowledge	3.414	3.210	3.490	0.204	-0.076	-0.280**
	(0.095)	(0.099)	(0.099)			
finance knowledge	2.909	2.690	2.960	0.219	-0.051	-0.270
	(0.122)	(0.121)	(0.115)			
risk preference	9.990	9.590	9.650	0.400	0.340	-0.060
	(0.225)	(0.222)	(0.222)			
Observations	99	100	100			

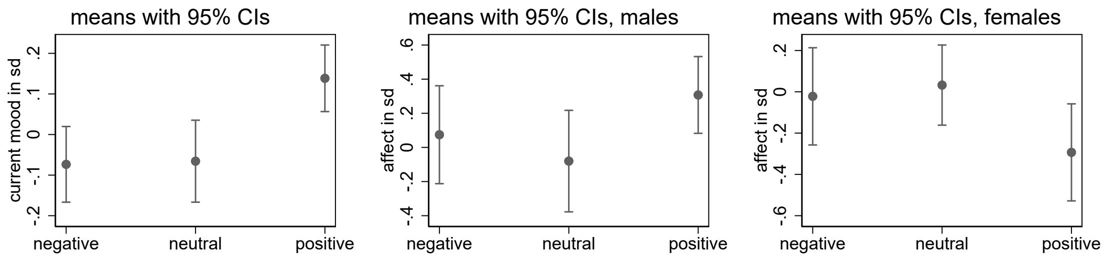
Notes: political preference: percentage of people who would vote for a certain party if there were national elections on the next Sunday; phone use: percentage of participants answering the survey on their phone; ex-ante feeling: self-evaluation on 11-point Likert scale; econ knowledge: self-evaluation about knowledge about the economy on 5-point Likert scale; finance knowledge: self-evaluation about knowledge about finance on 5-point Likert scale risk preference: self-evaluation on 11-point Likert scale.
The value displayed for t-tests in column 4,5 and 6 are the differences in the means across the respective groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

behavior. Moreover, the treatment effects vary by gender, suggesting that male and female participants respond differently to the emotional tone of headlines.

Current mood Participants who viewed the positive headline reported a significantly better current mood than those exposed to the neutral or negative versions, with an average difference of 0.42 points on the 11-point Likert scale ($p = 0.001$). Normalizing the outcome corresponds to an improvement of approximately 0.2 standard deviations. This effect is illustrated in Figure A25a. Regression results for both the general mood question and the affect scale are reported in Table A48.

Figure A25: Emotional Reactions by Readers

(a) average mood **(b)** average affect among males **(c)** average affect among females



Notes: Figure A25a illustrates means of the mood in the different treatment groups expressed in standard deviations. Figures A25b and A25c provide the means of aggregate affect in standard deviations by gender.

Affect No overall effect of headline sentiment on aggregate affect is detected, which appears to result from heterogeneous responses by gender, as shown in Figures A25b and A25c. Participants identifying as male report feeling more determined after reading the negative headline and more attentive, less afraid, and less upset after reading the positive one (each relative to the neutral condition). Hence, the positive headline appears to have an activating and positive effect on men's affect. In contrast, participants identifying as female report feeling less active when exposed to either the negative or the positive headline and less determined and alert after reading the positive one. As a result, the positive headline has an overall negative effect on women's affect, causing the aggregate effects to cancel out when both genders are considered jointly. These gender-specific responses are discussed in greater detail in Section A.3.2.4.

Directly related expectations The article presented to participants reported forecasts of German GDP growth, making their expectations about future GDP growth the most direct measure of how the provided information influenced belief updating. The outcome variable is defined as the absolute distance between each participant's forecast and the actual forecast stated in the article. Because participants could enter any numerical value, the data are susceptible to outliers; therefore, the variable is winsorized at the 95th percentile in the main specification. Similar results are obtained when winsorizing at the 99th percentile. For ease of interpretation, results are expressed in standard deviations.

The results indicate that the distance to the forecast is larger for participants exposed to emotional headlines, although only the estimate for 2022 is statistically significant. In this case, expectations of participants who viewed an emotional headline deviated by 0.23 standard deviations ($p = 0.047$) more from the actual forecast

Table A48: OLS Estimates – ATE on Emotions

	Current Mood in SD		Aggregate Affect in SD	
	(1)	(2)	(3)	(4)
positive	0.2175*	0.2029***	0.0428	-0.0067
	(0.1389)	(0.0655)	(0.1403)	(0.1137)
negative	-0.0390	-0.0023	0.0276	0.0361
	(0.1440)	(0.0702)	(0.1450)	(0.1249)
pre-feeling		0.4204***		0.2500***
		(0.0175)		(0.0271)
age		0.0024		0.0043
		(0.0042)		(0.0076)
phone		0.0027		0.1442
		(0.0891)		(0.1595)
male		-0.0293		-0.0070
		(0.0647)		(0.1171)
know econ		0.0161		0.0905
		(0.0427)		(0.0753)
know finance		0.0302		0.0176
		(0.0415)		(0.0608)
politics FE	no	yes	no	yes
Constant	-0.0598	-5.5289***	-0.0235	-3.8670***
	(0.1007)	(0.2777)	(0.1021)	(0.4738)
R^2	0.0128	0.7901	0.0003	0.3109
Observations	299	299	299	299

Notes: Table A48 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Pre-feelings are the answers to the current mode question prior to randomization on an 11-point Likert scale. Age is scaled in years. Phone is a dummy for survey answering on a smartphone. Know econ and know finance are answers to the self-evaluation questions on economic and financial knowledge on a 5-point Likert scale. *** p<0.01, ** p<0.05, * p<0.1

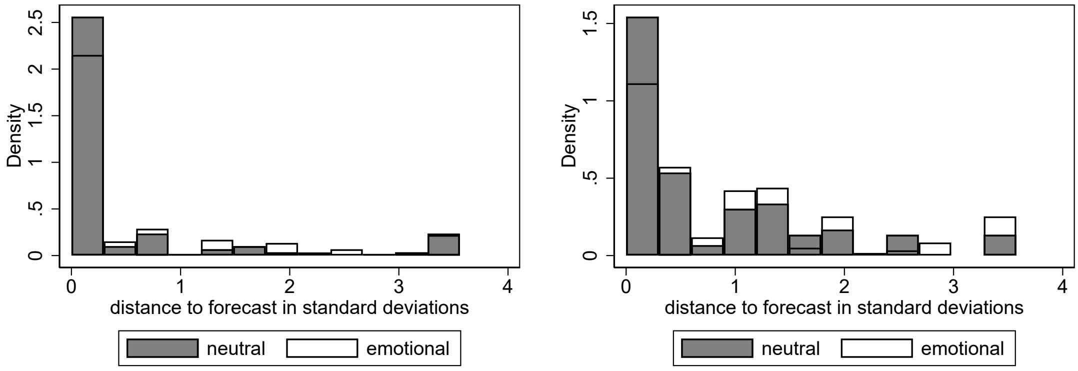
than those of participants exposed to the neutral headline. Figure A26 illustrates the distribution of forecast deviations, and full regression results are reported in Table A49.

Indirectly related expectations To examine whether readers extrapolate from the headlines to related, but only indirectly mentioned topics, expectations for the future development of the German stock index (DAX) were also elicited. As the DAX reflects the performance of the 40 largest German companies, it is correlated with overall economic growth, although changes in GDP do not translate perfectly into movements in the index. To anchor responses, all participants were shown the official DAX value from the week preceding the experiment. Reported expectations are winsorized and standardized for comparability.

The results show that exposure to the negative headline significantly lowers

Figure A26: Belief Updating by Readers

(a) 2021 **(b)** 2022



Notes: Figure A26 illustrates distribution of the distance to the forecasts expressed in standard deviations. The answers of the positive and negative group are pooled as "emotional" for better legibility. The answers of the neutral group are shaded in gray.

Table A49: OLS Estimates – ATE on GDP Expectations

	Expectations for 2021		Expectations for 2022	
	(1)	(2)	(3)	(4)
positive	0.1959	0.1720	0.1821	0.1745
	(0.1427)	(0.1399)	(0.1376)	(0.1435)
negative	0.1549	0.0930	0.2781**	0.2766*
	(0.1395)	(0.1449)	(0.1393)	(0.1427)
controls	no	yes	no	yes
Constant	-0.1170	06417	-0.1531*	0.2989
	(0.0977)	(1.0388)	(0.0908)	(0.9668)
R^2	0.0072	0.0615	0.0133	0.0442
Observations	299	299	299	299

Notes: Table A49 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Control variables are age, gender and political orientation of the participants as well as a dummy for phone-usage and self-reported measures for their feelings before the survey, knowledge on economics and knowledge on finance. *** p<0.01, ** p<0.05, * p<0.1

DAX expectations. Readers who viewed the negative headline expected the DAX to perform 0.27 standard deviations worse for 2021 ($p = 0.043$) and 0.29 standard deviations worse for 2022 ($p = 0.039$) compared to those exposed to the neutral headline. These effects correspond to a change in expectations of roughly 148 to 356 basis points, approximately equal to the average daily volatility of the DAX. Gender-specific analyses indicate a reaction only among men for the positive headline in 2022: male participants expected the DAX to perform 0.38 standard deviations worse ($p = 0.043$), whereas female participants, if anything, expected a slightly better performance. Full regression results for the DAX expectations are reported in Table A50.

Table A50: OLS Estimates – ATE on DAX Expectations

	Expectations for 2021		Expectations for 2022	
	(1)	(2)	(3)	(4)
positive	-0.1212	-0.1085	-0.1276	-0.1358
	(0.1483)	(0.1463)	(0.1482)	(0.1473)
negative	-0.3101**	-0.2729**	-0.2703**	-0.2931**
	(0.1369)	(0.1345)	(.1373)	(0.1413)
controls	no	yes	no	yes
Constant	0.1432	0.8887*	0.1322	0.1802
	(0.1069)	(0.5093)	(0.1065)	(0.4890)
R^2	0.0165	0.0648	0.0124	0.0628
Observations	299	299	299	299

Notes: Table A50 reports OLS estimates with robust standard errors in parentheses. All outcomes are scaled in standard deviations. The group that was exposed to the neutral headline is always the reference group. Control variables are age, gender and political orientation of the participants as well as a dummy for phone-usage and self-reported measures for their feelings before the survey, knowledge on economics and knowledge on finance. *** p<0.01, ** p<0.05, * p<0.1

Investment decisions No statistically significant effects of headline sentiment are detected in the incentivized investment task. Participants' investment choices remain unaffected by exposure to either positive or negative headlines.

Click rates Click rates do not differ significantly across headline types, with the exception of a marginally significant difference between the positive and neutral headlines ($p = 0.0626$ from a t-test). This difference, however, disappears once the standard set of covariates is included in the regression analysis.

A.3.2.4 Heterogeneous effects by gender

Readers' reactions to the headlines differ systematically by gender. The distortions in belief updating for GDP expectations and the more pessimistic DAX expectations following exposure to the negative headline are primarily driven by female participants. In contrast, the only statistically significant reaction among male participants occurs for the positive headline: men expect the DAX to perform worse in 2022 after reading it.[17]

A plausible explanation for these heterogeneous reactions may lie in the topic chosen for the experiment. Male and female participants may differ systematically

[17] One possible explanation for this seemingly counterintuitive reaction is that male participants may interpret the positive headline as a sign of current market overreaction and anticipate a subsequent correction in 2022.

in their prior attitudes toward the economy and financial markets. Consistent with this interpretation, Henkel and Zimpelmann, 2022 provide evidence that women, on average, view the stock market more negatively than men. If similar differences exist in the present sample and if participants exhibit confirmation bias, female participants may respond more strongly to the negative headline, because it may align more closely with their prior beliefs. In contrast, the positive headline appears to induce negative affect among women, particularly lower reported levels of activity, determination, and alertness, which may reflect a mismatch with their priors and consequently lead to limited belief updating. Overall, the observed heterogeneity highlights the importance of examining readers' responses to emotional headlines across a broader range of topics and contexts.

A.3.2.5 Limitations

The reactions of readers to the headlines provide initial evidence on potential economic effects of more emotional news coverage. However, this part of the experiment is likely underpowered, and the results should therefore be interpreted with caution. Statistical power is limited in part because not all participants chose to click on the article, meaning that the estimated effects on forecast accuracy should be viewed as intention-to-treat effects. True effects may be larger, and smaller effects may remain undetected. Since clicking behavior is potentially endogenous, it is not included as a control variable in the regression analysis. Moreover, it cannot be instrumented using treatment assignment, as treatment assignment does not predict clicking decisions.

Note that the absence of a treatment effect on clicking rates should not be interpreted as evidence that headline tone is irrelevant for reader engagement or that journalists are poor predictors of audience behavior. The clicking task in this setting primarily captures information-seeking behavior and does not fully reflect real-world online news consumption. Field evidence shows that emotional, and in particular negative, headlines are associated with higher click rates.

A.4 Experimental Instructions (English translation)

A.4.1 Experimental Instructions Journalist Experiment 1 (Pilot)

Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.

Welcome! Participation in this study takes **five minutes**. You can earn up to 15 Euro (but at least 5 Euro). The money will be transferred via PayPal within two weeks. All information provided here will be stored anonymously and used exclusively for research purposes. We will ask for your **PayPal address** at the end of the survey to be able to remunerate you. This data will be stored separately from the rest of your information and will be deleted after the payment is completed.

Are you currently working as a journalist?

O yes
O no

Participants who answered "no" had to leave the survey.

....

This page was only shown to the treatment group.

On the next page you will see a news article and three headlines. Your task is to choose the headline that you would most like to put above this article. The selected headline will be shown to a larger group of other participants (non-journalists). The more of those readers click on your selected headline, the more you will be paid.

Attention: The amount of your payout in this study depends on how many readers click on the headline you have selected.

....

This page was only shown to the control group.

On the next page you will see a news article and three headlines. Your task is to choose the headline that you would most like to put above this article. The selected headline will be shown to a larger group of other participants (non-journalists).

Which of the following headlines would you most likely put above the article below?

O **Encouraging forecast: the German economy is expected to grow strongly again in 2022**
O **Forecast: This is how the German economy will develop in the near future**
O **Scary forecast: 2021 will be worse than expected for the German economy**

In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.

The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.

Suppose you were allowed to formulate a headline yourself: What headline would you write?

Participants could type their answer into a free text field. They had the option to see the text of the article again through clicking on a button.

What aspects do you respect when choosing headlines for a given text? (multiple answers are possible)

O length of the headline
O comprehensibility
O active language (instead of passive expressions)
O usage of verbs
O usage of positive words
O usage of negative words
O usage of emotional words
O factual correctness

O most accurate description of the content of the text
O sparks curiosity
O topicality
O high degree of informativeness
O other

When participants checked the box "other" an additional field opened in which they could type in a free text answer in order to specify their response.

Which medium do you (mainly) work for?

Participants could type their answer into a free text field.

What is your main role in your journalistic work?
O writing/producing content
O edit content
O production of pictures, graphics and animations
O other

How many years have you been working in journalism?

Participants could type their answer into a free text field but the answer type was restricted to be a positive number.

How old are you?

Participants could type their answer into a free text field but the answer type was restricted to be a positive number.

What is your highest educational degree?

O no degree
O lower secondary school diploma
O higher secondary school diploma
O high-school diploma
O academic degree
O other

There are two types of secondary schools in the German school system which refer to different abilities of the students, i.e. obtaining a degree from a lower secondary school ("Hauptschule") is easier than obtaining one from a higher seconary school ("Realschule").

If there were federal elections next Sunday, which party would you vote for?

O CDU/CSU
O SPD
O Bündnis 90/Die Grünen
O FDP
O AfD
O Die Linke
O other
O I wouldn't vote.

In general, how willing or unwilling are you to take risks?

Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".

How willing would you be to give up something that benefits you today in order to benefit more in the future?

Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".

How willing would you be to give to a good cause without expecting anything in return?

Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".

How much do you agree with the statement "I am narcissistic."? *(Note: the word "narcissistic" means selfish, self-focused, and vain)*

Participants could answer the question on a 7-point Likert scale ranging form "don't agree at all" to "agree completly".

How much do you agree with the statement "I suspect people have only the best of intentions."?

Participants could answer the question on a 11-point Likert scale ranging form "don't agree at all" to "agree completly".

This page was only shown to the treatment group.

Thank you!

By participating, you have earned 5 Euro plus an additional amount depending on the clicks on the headline you selected. In order to be able to pay you, we need your PayPal address. It will be stored separately and deleted after your remuneration. The payments will be made within the next few days.

This page was only shown to the control group.

Thank you!

By participating, you have earned 10 Euro. In order to be able to pay you, we need your PayPal address. It will be stored separately and deleted after your remuneration. The payments will be made within the next few days.

A.4.2 Experimental Instructions Journalist Experiment 2

Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.

Welcome! In this survey, we aim to learn how journalists make content decisions when selecting headlines. Answering all questions will take about **10 minutes** and also works on a mobile phone (e.g., on the go). You can receive a participation fee of **10 Euro** plus an **additional payment that depends on your decisions**.

*This text was shown only to the **pay-per-click** and **pay-per-subscription** treatment groups. The wording of the compensation reflects the performance-dependent payment in these conditions.*

*Alternative version shown only to the **flat-pay** treatment group:*

Welcome! In this survey, we aim to learn how journalists make content decisions when selecting headlines. Answering all questions will take about **10 minutes** and also works on a mobile phone (e.g., on the go). You can receive a **flat participation fee of 15 Euro**.

The following text was displayed to participants in all groups:

This study is conducted by the Chair of Media Economics at the University of Cologne. All information provided here will be stored anonymously and used exclusively for research purposes. You may withdraw from the survey at any time without giving reasons. If you have questions or would like to exercise your rights under the General Data Protection Regulation (GDPR), please contact Lara Berger. Email: `lara.berger@uni-koeln.de`

Details on Data Processing and Data Protection

To process your payment, we will ask for your **bank details** at the end of the survey. For legal reasons, this also includes your **name and address**. This information will be stored separately from all other responses, used only by the University of Cologne for payment purposes, and deleted after payment has been completed.

The heading "Details on Data Processing and Data Protection" was displayed as a *clickable button. If participants clicked on it, the following text was shown:*

For research purposes, we collect your evaluations of news headlines and other media content. Your responses will be processed exclusively within this research project. The data will not be shared with third parties. Your personal data will not be used for automated decision-making, including profiling, in accordance with Art. 22 (1) and (4) GDPR.
The legal basis for data processing is your consent (Art. 6 (1) (a) GDPR). Participation in this research project is voluntary. Providing your personal data is neither legally nor contractually required and not necessary for the conclusion of a contract. If you decline to participate or withdraw your consent, you will not experience any disadvantages. Withdrawal of consent only affects future processing; data processing that took place prior to withdrawal remains lawful.
Under the General Data Protection Regulation (GDPR), you have the following rights: (1) You may request information about whether and which personal data we process and obtain related information (Art. 15 GDPR). Please note that this right may be restricted or excluded in certain cases (cf. Art. 10 BayDSG). (2) If inaccurate personal data are processed, you have the right to rectification (Art. 16 GDPR). (3) Provided legal conditions are met, you may request deletion or restriction of processing of your personal data (Art. 17 and 18 GDPR). However, the right to deletion under Art. 17 (1–2) GDPR does not apply when processing is necessary to perform a task in the public interest or under official authority (Art. 17 (3) (b) GDPR) or if legal retention obligations exist. (4) You have the right to receive the data you provided in a standardized electronic format or have it transmitted to another entity, if the processing is based on consent or contract and carried out by automated means (Art. 20 GDPR). (5) You have the right to withdraw your consent at any time with effect for the future. Withdrawal does not affect the lawfulness of processing carried out prior to withdrawal. Please note that even after withdrawal, data processing may continue on another legal basis (cf. Art. 17 (1) (b) and (3) (d) GDPR). (6) You have the right to lodge a complaint with a supervisory authority pursuant to Art. 51 GDPR concerning the processing of your personal data. The responsible authority for the University of Cologne is the State Commissioner for Data Protection and Freedom of Information of North Rhine-Westphalia, reachable at Kavalleriestraße 2–4, 40213 Düsseldorf, or via `https://www.ldi.nrw.de/kontakt/ihre-beschwerde`. Before submitting a complaint, please contact the responsible unit conducting the research project, as this allows the matter to be resolved most efficiently.
If you wish to exercise one of these rights or have questions, you may contact the responsible research unit at any time using the contact details provided. We will

promptly review whether the legal requirements are met and take the necessary steps. Further restrictions, modifications, or exclusions of these rights may apply under the GDPR or national law—e.g., where exercising these rights would likely render the research objectives impossible or seriously impair them, and such restriction is necessary for achieving the research purposes (cf. Art. 25 (4) BayDSG).

Consent Form

I have received and read the information sheet about the research project and consent to participate in this project and to the associated processing of my data. I understand that I may decline participation without any disadvantages.

O I consent
O I do not consent

Participants who selected "I do not consent" were redirected to the end of the survey and could not continue.

Are you currently working as a full-time journalist and/or have you worked as a journalist in the past 12 months?

O yes
O no

Participants who answered "no" were shown the following message and the survey ended.

Thank you for your interest! Unfortunately, you do not meet the eligibility criteria for participation in this study.

If you have any questions, please contact `lara.berger@uni-koeln.de`.

The following page contained task instructions. The text varied depending on the experimental group.

Version shown to the ***flat-pay*** *group:*

In the following, you will be shown two short articles.

Your task is to select headlines from several options and to propose your own headline for each text. We previously showed the available headline options to a representative sample of readers from the German population. The headlines you propose will later be shown to a smaller group of readers.

*Version shown to the **pay-per-click** group:*

In the following, you will be shown two short articles.

Your task is to select headlines from several options and to propose your own headline for each text. Each headline you select affects your payment. Your total remuneration at the end of the survey consists of **6 Euro plus the amount you earn through your headline choices**. You will receive at least **13 Euro in total**.

We previously showed the available headline options to a representative sample of readers from the German population and recorded how often they clicked on each headline. For every headline you choose, you will earn **4 Euro multiplied by the share of readers who clicked on it**.

Example: If 50 percent of readers clicked on the headline you selected, you will receive 2 Euro for that headline. If only 10 percent clicked on the next one, you will receive 40 cents for that headline.

Thus, the more clicks your selected headlines generate, the higher your payment will be. At the end, all amounts will be added up and paid out.

The headlines you write yourself can also increase your payout. After the survey, these self-written headlines will be shown to a group of readers to measure how often they click on them. The authors of the five headlines that receive the most clicks will each receive an **additional payment of 50 Euro**.

In other words: the more clicks your own headline generates, the higher your chance of receiving an additional 50 Euro.

*Version shown to the **pay-per-subscription** group:*

In the following, you will be shown two short articles.

Your task is to select headlines from several options and to propose your own headline for each text. Each headline you select affects your payment. Your total remuneration at the end of the survey consists of **6 Euro plus the amount you earn through your headline choices**. You will receive at least **10 Euro in total**.

We previously showed the available headline options to a representative sample of readers from the German population. Afterwards, these readers were given the opportunity to **subscribe to a free newsletter** if they wanted to learn more about the topic. For every headline you choose, you will earn **8 Euro multiplied by the share of readers who subscribed**.

Example: If 30 percent of readers subscribed after reading your headline, you will receive 2.40 Euro for that headline. If only 10 percent subscribed after reading the next one, you will receive 80 cents.

Thus, the more newsletter subscriptions your selected headlines generate, the higher your payment will be. At the end, all amounts will be added up and paid out.

The headlines you write yourself can also increase your payout. After the survey, these self-written headlines will be shown to a group of readers to measure how often they would subscribe to a free newsletter on the same topic. The authors of the five headlines that generate the most subscriptions will each receive an **additional payment of 50 Euro**.

In other words: the more subscriptions your own headline generates, the higher your chance of receiving an additional 50 Euro.

Please read the following news story:

—first paragraph of one randomly drawn article from the article pool—

Participants could click a button labeled "more" to display additional paragraphs of the article.

Which headline would you suggest for the news story above?

The order of the following tasks (competition vs. no competition) was randomized across participants.

Which of the following three headlines would you most likely place above the news story?

– option existing 1
– option existing 2
– option existing 3

Show article again
Clicking this button displayed the article content again.

Note: The headlines presented here were shown to readers **individually**. This means that different reader groups were each exposed to only one of these headlines.

Which of the following three headlines would you most likely place above the news story?

– option gpt 1
– option gpt 2
– option gpt 3

Show article again
Clicking this button displayed the article content again.

Note: The headlines presented here were shown to readers **individually**. This means that different reader groups were each exposed to only one of these headlines.

Competition condition.

Note: The headlines shown in the following task were presented to readers **side by**

side with another headline. Readers could choose to read the text associated with one, both, or neither of the two headlines. The following headline competed with the one you selected earlier:
"text competitive headline"

Which of the following three headlines would you most likely place above the news story?

– option existing 1
– option existing 2
– option existing 3

Show article again
Clicking this button displayed the article content again.

Competition condition continued.

Note: The headlines shown in the following task were presented to readers **side by side with another headline**. Readers could choose to read the text associated with one, both, or neither of the two headlines. The following headline competed with the one you selected earlier:
"text competitive headline"

Which of the following three headlines would you most likely place above the news story?

– option gpt 1
– option gpt 2
– option gpt 3

Show article again
Clicking this button displayed the article content again.

Please read the following news story:

—first paragraph of one randomly drawn article from the article pool—

Participants could click a button labeled "more" to display additional paragraphs of the article.

Which headline would you suggest for the news story above?

The order of the following tasks (competition vs. no competition) was randomized across participants.

Which of the following three headlines would you most likely place above the news story?

– option existing 1
– option existing 2
– option existing 3

Show article again
Clicking this button displayed the article content again.

Note: The headlines presented here were shown to readers **individually**. This means that different reader groups were each exposed to only one of these headlines.

Which of the following three headlines would you most likely place above the news story?

– option gpt 1
– option gpt 2
– option gpt 3

Show article again
Clicking this button displayed the article content again.

Note: The headlines presented here were shown to readers **individually**. This means that different reader groups were each exposed to only one of these headlines.

Competition condition.

Note: The headlines shown in the following task were presented to readers **side by side with another headline**. Readers could choose to read the text associated with one, both, or neither of the two headlines. The following headline competed with the one you selected earlier:
"text competitive headline"

Which of the following three headlines would you most likely place above the news story?

– option existing 1
– option existing 2
– option existing 3

Show article again
Clicking this button displayed the article content again.

Competition condition continued.

Note: The headlines shown in the following task were presented to readers **side by side with another headline**. Readers could choose to read the text associated with one, both, or neither of the two headlines. The following headline competed with the one you selected earlier:
"text competitive headline"

Which of the following three headlines would you most likely place above the news story?

– option gpt 1
– option gpt 2
– option gpt 3

Show article again
Clicking this button displayed the article content again.

Thank you! You have now almost reached the end of the survey.

Finally, we have a few questions about you.

Small note displayed below the question block:
We use these answers to analyze the data in relation to certain demographic patterns — for example, whether women behave systematically differently than men.

How old are you?

Open text field (only positive integers accepted).

With which gender do you identify?

– male
– female
– diverse
– prefer not to say

Where is your main place of residence?

– Western Germany
– Eastern Germany
– Outside Germany
– prefer not to say

What is your highest educational qualification?

– University or university of applied sciences degree
– High school diploma (Abitur or Fachabitur)
– Secondary school diploma (Realschule)
– Lower secondary school diploma (Hauptschule)

– No school diploma
– prefer not to say

If the federal election were held next Sunday, which party would you vote for?

– CDU/CSU
– SPD
– The Greens
– FDP
– AfD
– The Left (Die Linke)
– Other
– None / I would not vote
– prefer not to say

What type of medium do you usually work for?
If you work for several types of media, please select the category for which you work most frequently.

– Print: newspapers, magazines, or trade journals
– Online: newspapers, magazines, or trade journals
– Other online media (social media, blogs, etc.)
– Private broadcasting
– Public broadcasting
– Other
– prefer not to say

If "Other" was selected, an open text field appeared allowing participants to specify their answer.

Which role do you usually take on in your journalistic work?
If you regularly take on more than one role, you could select multiple options.

– Reporter

– Editor
– Fact-checker / Researcher
– Proofreader
– Correspondent
– Columnist
– Investigative journalist
– Other

If "Other" was selected, an open text field appeared allowing participants to specify their answer.

If you still have time: could you tell us according to which criteria headlines are usually selected for the articles you produce?

If you would like to finish quickly, simply click the arrow in the bottom right corner.

Large open text field for free responses.

Finally, we would like to ask for your feedback on this study.
Your feedback helps us improve our research.

For example: Were there any sections of the study that you did not fully understand, or where you would have liked more detailed explanations?

Again, if you would like to finish quickly, simply click the arrow in the bottom right corner.

Large open text field for free responses.

Final page shown only to participants in the ***pay-per-click*** *treatment group.*

Thank you! You have now reached the end of the survey.

If you wish, you can click the following button to see how often the headlines you selected were clicked on:

Detailed click data (button)

For your headline decisions, you earned a total of $e://Field/bonus_total Euro.

Your total remuneration therefore amounts to $e://Field/payoff Euro.

If one of your self-written headlines ranks among the five headlines with the highest number of clicks, you will receive an additional payment of 50 Euro. These funds will be paid out separately from your current remuneration. Please note that it may take several weeks before these additional payments are made.

Final page shown only to participants in the ***pay-per-subscription (pay-for-abo)*** *treatment group.*

Thank you! You have now reached the end of the survey.

If you wish, you can click the following button to see how often readers subscribed to a newsletter after reading the headlines you selected:

Detailed subscription data (button)

For your headline decisions, you earned a total of $e://Field/bonus_total Euro.

Your total remuneration therefore amounts to $e://Field/payoff Euro.

If one of your self-written headlines ranks among those with the highest number of subscriptions, you will receive an additional payment of 50 Euro. These funds will be paid out separately from your current remuneration. Please note that it may take several weeks before these additional payments are made.

Final page shown only to participants in the ***flat-pay*** *treatment group.*

Thank you! You have now reached the end of the survey.

As a participation fee, you will receive 15 Euro.

Payment procedure shown to all participants.

You will now be redirected to the **payment form of the Cologne Laboratory for Economic Research (CLER)**. Please do not be surprised: the wording used there is primarily intended for students who regularly take part in studies conducted by the laboratory.

For legal reasons, we also use this form for the present study. Your payment details will thus be stored securely and separately from your survey responses, and they will be deleted immediately after payment is completed.

Your payment code is `-- individualized code --`. You do not need to note this code—it will be transmitted automatically to the CLER form when you click the arrow below.

Please note that payment may take up to two weeks after your participation in the study.

A.4.3 Experimental Instructions Readers Experiment 1 (Pilot)

Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.

Welcome!

At the **end of this study** you will be given a **withdrawal code**, please write it down. After your participation we will redirect you to the withdrawal form. You may have to enter this code there.

This participation in this study will take approximately 5 minutes.

O I consent to the above conditions.

. . . .

In general, how do you feel right now?

Participants could answer the question on a 11-point Likert scale ranging form "very bad" to "very good".

. . . .

How good or bad is your knowledge on current topics in the field of economics and business?

O very bad
O bad
O neither good nor bad
O good
O very good

. . . .

How good or bad is your knowledge on current topics in the field of finance?

O very bad
O bad
O neither good nor bad
O good
O very good

In general, how willing or unwilling are you to take risks?

Participants could answer the question on a 11-point Likert scale ranging form "not willing at all" to "very willing".

On the next page we will show you the **headline** of a news article. **Please read this headline carefully.**

Attention: If you **click on the headline** you can read the **entire article**. **This will cost you 5 cents.** The contents of the article may help you make more informed decisions later in the study. You can earn up to one euro for the decisions you make later. Once you've clicked on the headline, you can open or close the article as many times as you want at no additional cost.

This page was only displayed to participants in the positive treatment group.

Encouraging forecast: the German economy is expected to grow strongly again in 2022

When participants clicked on the headline the entire article from the journalist experiment was displayed. Otherwise they only saw the headline.

In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.

The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.

This page was only displayed to participants in the neutral treatment group.

Forecast: This is how the German economy will develop in the near future

When participants clicked on the headline the entire article from the journalist exper-

iment was displayed. Otherwise they only saw the headline.

In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.

The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.

This page was only displayed to participants in the negative treatment group.

Scary forecast: 2021 will be worse than expected for the German economy

When participants clicked on the headline the entire article from the journalist experiment was displayed. Otherwise they only saw the headline.

In their autumn report, which published in October, leading economic research institutes expect gross domestic product to grow by 2.4 percent in 2021. In spring they had still expected an increase of 3.7 percent for this year.

The economic situation in Germany is still characterized by the corona pandemic, the report says. In the course of 2022, however, the German economy is expected to return to normal capacity. According to forecasts by the institutes, gross domestic product will increase by 4.8 percent in 2022. In their spring forecast, the institutes had only assumed an increase of 3.9 percent for the next year.

In general, how do you feel right now?

Participants could answer the question on a 11-point Likert scale ranging form "very bad" to "very good".

The following words describe different feelings and sensations. Read every word, then indicate the intensity with which you experience the respective emotion at the moment. You can choose between five gradations.

	not at all	a little	somewhat	much	very much
upset	○	○	○	○	○
attentive	○	○	○	○	○
afraid	○	○	○	○	○
determined	○	○	○	○	○
nervous	○	○	○	○	○
inspired	○	○	○	○	○
active	○	○	○	○	○
hostile	○	○	○	○	○
awake	○	○	○	○	○
ashamed	○	○	○	○	○

You will now make four estimates and two decisions. One of these answers will be selected at random and will become relevant for your remuneration.

The following applies to the following questions: The closer the estimated value to the actual realized value, the higher your payout. You can earn up to one euro.

Note: Your payout is independent of what other participants value. **You should therefore enter the value that you consider most probable for each guess question.**

Note: For the following two questions, enter your answer as a percentage. Example: If you expect a one percent increase, enter "1". If you expect a one percent drop, enter "-1".

What do you expect: by what percentage will the German gross domestic product increase or decrease in the course of **2021** compared to 2020?

Participants could answer the question in a free text field.

What do you expect: by what percentage will the German gross domestic product increase or decrease in the course of **2022** compared to 2021?

Participants could answer the question in a free text field.

Note: On December 3, 2021, the Dax closed at 15,169 points.

What do you expect: With how many points will the Dax close on December 31, 2021 (i.e. at the end of **this** year)?

Participants could answer the question in a free text field.

What do you expect: With how many points will the Dax close on December 31, 2022 (i.e. the end of **next** year)?

Participants could answer the question in a free text field.

You can now invest all or part of 50 cents in the DAX. The invested money remains invested **until December 31, 2021**.

Your money (i.e. both the uninvested and the invested money) will be paid out on January 1st, 2022. The amount of uninvested money remains the same. The amount of money invested depends on the development of the DAX.

Example: If the DAX rises by two percent and you invest 50 cents, you will be paid 51 cents. If it falls by two percent, you will receive 49 cents.

How much of the 50 cents do you want to invest?

Participants could answer the question in a free text field.

You can now invest all or part of 50 cents in the DAX. The money remains invested **until December 31, 2022**.

How much of the 50 cents do you want to invest?

Participants could answer the question in a free text field.

How old are you?

Participants could answer the question in a free text field, but answers were restricted to positive numeric values.

What gender do you feel you belong to?

O male
O female
O diverse

If there were federal elections next Sunday, which party would you vote for?

O CDU/CSU
O SPD
O Bündnis 90/Die Grünen
O FDP
O AfD
O Die Linke
O other
O I wouldn't vote.

Thanks!

Your personal payout code is: XXX.

Please click on the arrow below to be redirected to the withdrawal form.

A.4.4 Experimental Instructions Readers Experiment 2

Participants received experimental instructions in German. Below I provide an English transcript. A dotted line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.

Welcome!

In this survey, we would like to learn more about your interest in different news articles and topics that are frequently covered in the media.

All information provided here will be stored anonymously and used exclusively for research purposes. You may withdraw from the survey at any time and without providing a reason. However, to ensure that your responses can be used for research purposes, it is important that you answer all questions completely.

Detailed information on data protection is provided on the following page.

If you have any questions while completing the survey, please feel free to contact `lara.berger@uni-koeln.de`.

. . . .

Information about the survey and data protection

1. General information

The aim of this survey is to understand how much interest different population groups show in various news articles and topics that are frequently covered in the media, and how this interest varies across different circumstances. Data are collected through an online questionnaire. The survey is conducted by researchers at the University of Cologne.

2. Responsible institution and contact details

The survey is carried out under the responsibility of the University of Cologne. All responses are stored in anonymized form.

If you have any questions about the project or about data processing, if you wish to withdraw your consent, or if you would like to exercise your rights under the General

Data Protection Regulation (GDPR), please contact Lara Berger.

E-mail: `lara.berger@uni-koeln.de`

Details on data processing and data protection

"Details on data processing and data protection" was a clickable button. This text appeared after clicking the button:

For research purposes, we collect information about your interest in different news articles and topics that are frequently covered in the media. Your responses will be processed exclusively within the scope of this research project and will not be shared with third parties. Your personal data will not be used for automated decision-making, including profiling, within the meaning of Article 22(1) and (4) of the GDPR.

The legal basis for data processing is **your consent** (Article 6(1)(a) GDPR). Participation in the research project is **voluntary**. Providing your personal data is neither legally nor contractually required, nor necessary for the conclusion of a contract. If you decline participation or withdraw your consent, this will have **no disadvantages** for you. A withdrawal of consent only applies to future processing; data processing that took place before the withdrawal remains lawful.

Under the GDPR, you have the following rights: (1) You may request information about whether and which personal data we process about you, as well as related information (Article 15 GDPR). Please note that this right of access may be limited or excluded in certain cases. (2) If inaccurate personal data are processed, you have the right to rectification (Article 16 GDPR). (3) Provided the legal requirements are met, you may request the deletion or restriction of the processing of your personal data (Articles 17 and 18 GDPR). The right to erasure under Article 17(1) and (2) GDPR does not apply, for example, when processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority (Article 17(3)(b) GDPR), or when legal retention obligations exist. (4) You have the right to receive data you have provided in a structured, commonly used, and machine-readable format or to have it transmitted to another controller, where technically feasible (Article 20 GDPR). (5) You have the right to withdraw your consent at any time with future effect. Withdrawal does not affect the lawfulness of processing based on consent before its withdrawal. Please note that further processing may still be possible on other legal grounds (see Article 17(1)(b) and 17(3)(d) GDPR). (6) You have the right to lodge a complaint with a supervisory authority within the meaning of Article 51 GDPR regarding the processing of your personal data. The competent supervisory authority for the University of Cologne is the State Commissioner for Data Protection and Freedom of Information of North Rhine-Westphalia, reachable at Kavalleriestraße 2–4, 40213 Düsseldorf, or via `https://www.ldi.nrw.de/kontakt/ihre-beschwerde`. Before submitting a complaint, you should contact the responsible organizational unit (see Section 2) to ensure your concern can be addressed as quickly as possible.

If you wish to exercise any of these rights or have further questions, please contact the research team using the details provided above. We will promptly review whether the legal requirements are met and take the necessary measures. Further limitations or exclusions of these rights may arise from the GDPR or national legislation, for example if exercising these rights would make it impossible or seriously impair the achievement of the research purposes and if such limitations are necessary to fulfill those research purposes. (See Article 25(4) BayDSG.)

Declaration of consent

I have received and read the information sheet about the research project. I consent to participate in this research project and to the associated processing of my personal data. I am aware that I may refuse to participate and that non-participation will not result in any disadvantages for me. I have been explicitly informed about the possibility of withdrawing my consent at any time.

O I consent
O I do not consent

Participants who did not provide informed consent were excluded from further participation.

In general, how do you feel right now?

On the scale, 0 means “very bad” and 100 means “very good.”

Participants indicated their answer using a slider ranging from 0 to 100.

In the following, we ask you to assess your prior knowledge in different subject areas.

Your answers will neither affect the further course of the survey nor your compensation.

There are no right or wrong answers.

How good or bad is your knowledge of current topics in the field of European politics?

O very bad
O bad
O neither good nor bad
O good
O very good

How good or bad is your knowledge of current topics in the field of technology (in particular artificial intelligence)?

O very bad
O bad
O neither good nor bad
O good
O very good

How good or bad is your knowledge of current topics in the field of space missions?

O very bad
O bad
O neither good nor bad
O good
O very good

How good or bad is your knowledge of current topics in the field of economics (in particular inflation)?

O very bad
O bad
O neither good nor bad
O good
O very good

How good or bad is your knowledge of current topics in the field of crime?

O very bad
O bad
O neither good nor bad
O good
O very good

How good or bad is your knowledge of current topics in the field of consumption (in particular products for children)?

O very bad
O bad
O neither good nor bad
O good
O very good

In surveys like this one, it unfortunately often happens that participants do not read the instructions and questions carefully and simply click through quickly. As a result, many questions are answered randomly, which negatively affects the survey results. To show that you read the instructions and questions thoroughly, please answer the following question with “green.”

Based on what you just read, what is your favorite color?

O red
O purple
O green
O yellow

O blue

Only participants who passed this attention check were able to proceed with the survey.

We are now interested in your opinion on different topics.

You will be presented with several statements, and you can indicate the extent to which you agree or disagree with each statement.

There are no right or wrong answers.

To what extent do you agree with the following statement:

"EU regulations for products and services increase consumer prices and harm small businesses."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Without strict regulation, artificial intelligence will destabilize society."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Money spent on space missions is a waste, we already have enough problems on

Earth."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Rapid interest rate cuts are risky and should be avoided."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Social problems are the main cause of rising crime rates."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"When it comes to child car seats, safety should always be more important than price."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.

O I somewhat agree with the statement.
O I completely agree with the statement.

Thank you! We now have a few questions about yourself.

What is your highest level of education?

O University degree or university of applied sciences degree
O High school diploma (Abitur) or advanced technical college entrance qualification
O Secondary school diploma (Realschule)
O Lower secondary school diploma (Hauptschule)
O No school-leaving qualification

If federal elections were held next Sunday, which party would you vote for?

O CDU/CSU
O SPD
O The Greens
O FDP
O AfD
O The Left (Die Linke)
O Other
O None / I would not vote
O Prefer not to say

Which activities best describe your current occupation?

O Management activities
O Commercial or administrative activities
O Food production and processing
O Tourism, hotel, or catering professions
O Building and property cleaning
O IT, mathematics, or natural sciences
O Humanities and social sciences

O Personal care and wellness
O Architecture and engineering
O Sales and retail
O Education, training, or library work
O Legal professions
O Social work and related activities
O Agriculture, fisheries, or forestry
O Art and design
O Advertising, marketing, and media
O Performing and entertainment professions
O Sports and sports management
O Construction and extraction
O Medicine and health
O Manufacturing and production
O Protection and security
O Transport and logistics
O Other

What is your household's monthly net income (after taxes)?

O less than €1,500
O €1,500 – €2,500
O €2,500 – €3,500
O €3,500 – €4,500
O €4,500 – €5,500
O more than €5,500
O Prefer not to say

How often do you consume printed news content (e.g., newspapers or magazines)?

O Daily, more than two hours
O Daily, one to two hours
O Daily, but less than one hour
O Several times per week
O Several times per month
O Less than once per month

O Never

How often do you consume digital news content (e.g., via news websites, news videos, or news content on social media)?

O Daily, more than two hours
O Daily, one to two hours
O Daily, but less than one hour
O Several times per week
O Several times per month
O Less than once per month
O Never

In surveys like this one, it unfortunately often happens that participants do not read the instructions and questions carefully and simply click through quickly. As a result, many questions are answered randomly, which negatively affects the survey results. To show that you read the instructions and questions thoroughly, please answer the following question with "horse."

Based on what you just read, what is your favorite animal?

O Dog
O Cat
O Mouse
O Horse
O Bird

Participants who did not pass this attention check were excluded from further participation.

This page was only displayed to participants in the no-competition treatment.

Next, we will present you with a series of different headlines.

If you wish, you can click on a headline to read the corresponding article.

If you are interested in the topic, you can subscribe to a newsletter in which you will receive a one-time compilation of additional articles on the topic.

Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once for sending the newsletter, and deleted immediately after the newsletter has been sent.

This page was only displayed to participants in the no-competition treatment.

In this section, we are interested in your interest in different articles from news media. The articles presented to you are based on real news items.

On the following pages, please behave as you would if you encountered one of these headlines elsewhere on the internet, and click on it only if you would probably do so outside the context of this survey.

This page was only displayed to participants in the competition treatments.

Next, we will present you with pairs of two different headlines.

If you wish, you can click on one or both of the headlines to read the corresponding articles.

If you are interested in the topic, you can subscribe to a newsletter in which you will receive a one-time compilation of additional articles on the topic.

Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once for sending the newsletter, and deleted immediately after the newsletter has been sent.

This page was only displayed to participants in the competition treatments.

In this section, we are interested in your interest in different articles from news media. The articles presented to you are based on real news items.

On the following pages, please behave as you would if you encountered one of these headlines elsewhere on the internet, and click on it only if you would probably do so outside the context of this survey.

Subsequently, one of the headlines from the topic pool was randomly drawn and was displayed with valence according to the participant's treatment group. Participants in the neutral/no-competition treatment saw a neutral headline alone, while those in the neutral/competition group saw the same headline alongside the corresponding competing one. The same applied to the positive/competition, negative/competition, positive/no-competition, and negative/no-competition groups. Participants could click on a headline to read the corresponding article text. The time they spend with each headline and/or text was recorded. The complete pool of articles and all possible headlines is provided in Appendix A.2.1.1.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic *X*?

O Yes
O No

"X" was replaced by the respective article topic.
Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the newsletter has been sent.

Subsequently, a corresponding headline from another randomly drawn topic was displayed. Earlier drawn topics were excluded from future randomization.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic *X*?

O Yes
O No

"X" was replaced by the respective article topic.
Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the newsletter has been sent.

Subsequently, a corresponding headline from another randomly drawn topic was displayed.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic X?

O Yes
O No

"X" was replaced by the respective article topic.
Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the newsletter has been sent.

Subsequently, a corresponding headline from another randomly drawn topic was displayed.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic X?

O Yes
O No

"X" was replaced by the respective article topic.

Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the newsletter has been sent.

Subsequently, a corresponding headline from another randomly drawn topic was displayed.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic *X*?

O Yes
O No

"X" was replaced by the respective article topic.

Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the newsletter has been sent.

Subsequently, a corresponding headline from another randomly drawn topic was displayed.

Would you like to receive a free e-mail newsletter with an overview of additional articles on the topic *X*?

O Yes
O No

"X" was replaced by the respective article topic.

Note on data processing: If you decide to subscribe to a newsletter, we will collect your e-mail address at the end of this survey. This address will be stored separately from your other responses, used only once to send the newsletter, and deleted immediately after the *newsletter has been sent.*

In general, how do you feel right now?

On the scale, 0 means “very bad” and 100 means “very good.”

Participants indicated their answer using a slider ranging from 0 to 100.

In the following, we ask you to answer a few estimation questions.

Your answers will neither affect the further course of the survey nor your compensation.

How many kebab shops are there in Germany?

Participants could answer the question in an open text field.

According to estimates by the International Monetary Fund, what percentage of jobs in Germany could be at risk due to artificial intelligence?

Participants could answer the question in an open text field.

How many kilometers away is the asteroid “Apophis” predicted to pass by Earth in the year 2029?

Participants could answer the question in an open text field.

What was the core inflation rate in the euro area in August 2024 (in percent)?

Participants could answer the question in an open text field.

According to crime statistics, how many criminal offenses were recorded in South-East Hesse in 2023?

Participants could answer the question in an open text field.

In 2024, the consumer organization *Stiftung Warentest* tested a total of 24 child car seats and baby carriers. How many of them received the rating "good"?

Participants could answer the question in an open text field.

In surveys like this one, it unfortunately often happens that participants do not read the instructions and questions carefully and simply click through quickly. As a result, many questions are answered randomly, which negatively affects the survey results. To show that you read the instructions and questions thoroughly, please answer the following question with "rose."

Based on what you just read, what is your favorite plant?

O Rose
O Palm tree
O Sunflower
O Cactus
O Daisy

Participants who did not pass this attention check were excluded from further participation.

We are now again interested in your opinion on different topics.

You will be presented with several statements, and you can indicate the extent to which you agree or disagree with each statement.

There are no right or wrong answers.

To what extent do you agree with the following statement:

"It is exaggerated that the EU intervenes in detailed matters such as food preparation, which should be regulated nationally instead."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Artificial intelligence will create more jobs than it will destroy."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Space missions are essential to protect the Earth."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Prices are currently too high and should go down again, no matter what."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Harsher punishments are necessary to combat crime."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

To what extent do you agree with the following statement:

"Expensive child car seats do not offer any real advantage over cheaper models."

O I completely disagree with the statement.
O I somewhat disagree with the statement.
O I neither agree nor disagree with the statement.
O I somewhat agree with the statement.
O I completely agree with the statement.

Thank you! You have now almost reached the end of the survey. Finally, we have a few concluding questions for you.

Did you search for additional information on the internet while answering the estimation questions?

O Always
O Often

O Sometimes
O Rarely
O Never

You indicated at one or more points that you would like to receive a newsletter by e-mail. Please enter your e-mail address in the field below.

Participants could answer the question in an open text field.

Note on data processing: Your e-mail address will be stored separately from your other responses, used only once to send the newsletter(s), and deleted immediately after the newsletter(s) has been sent.

We would now like to ask for your feedback on this study. Your input helps us improve our research. For example, were there any sections of this study that you did not understand or where you would have preferred more detailed explanations?

Participants could answer the question in an open text field.

Appendix B

Appendix to Chapter 3

B.1 Supplementary material

B.1.1 Ten tips to spot false news

1. **Be skeptical of headlines.** False news stories often have catchy headlines in all caps with exclamation points. If shocking claims in the headline sound unbelievable, they probably are.

2. **Look closely at the link.** A phony or look-alike link may be a warning sign of false news. Many false news sites mimic authentic news sources by making small changes to the link. You can go to the site to compare the link to established sources.

3. **Investigate the source.** Ensure that the story is written by a source that you trust with a reputation for accuracy. If the story comes from an unfamiliar organization, check their "About" section to learn more.

4. **Watch for unusual formatting.** Many false news sites have misspellings or awkward layouts. Read carefully if you see these signs.

5. **Consider the photos.** False news stories often contain manipulated images or videos. Sometimes the photo may be authentic, but taken out of context. You can search for the photo or image to verify where it came from.

6. **Inspect the dates.** False news stories may contain timelines that make no sense, or event dates that have been altered.

7. **Check the evidence.** Check the author's sources to confirm that they are accurate. Lack of evidence or reliance on unnamed experts may indicate a false news story.

8. **Look at other reports.** If no other news source is reporting the same story, it may indicate that the story is false. If the story is reported by multiple sources you trust, it's more likely to be true.

9. **Is the story a joke?** Sometimes false news stories can be hard to distinguish from humor or satire. Check whether the source is known for parody, and whether the story's details and tone suggest it may be just for fun.

10. **Some stories are intentionally false.** Think critically about the stories you read, and only share news that you know to be credible.

B.1.2 List experiments

This section displays translations of the statements that we used in our list experiments on Corona vaccination and dietary supplements in Wave I and Wave II of our survey, respectively. The statements were shown in randomized order. The statements in regular font were shown to every participant, the statements in bold font only to about 50% of them. Assignment to see the additional statement was random.

List experiment on **Corona vaccination, Wave I**:

- I do not eat meat.
- I like football.
- I listen to the news on the radio in the morning.
- I live in a relatively small town.
- I usually go to bed rather late.
- **I prefer not to get vaccinated against Covid-19.**

List experiment on **dietary supplements, Wave I**:

- I like to go dancing.
- I work part-time but would prefer to work more.
- I like winter time.
- I do not have any pets.
- I suffer from a pollen allergy.
- **I consume dietary supplements.**

List experiment on **Corona vaccination, Wave II**:

- I like to go for a walk.
- I drink a lot of coffee.
- I do not have any siblings.
- My apartment is on the first floor.

- I like to eat bananas.
- **I prefer not to get vaccinated against Covid-19.**

List experiment on **dietary supplements, Wave II**:

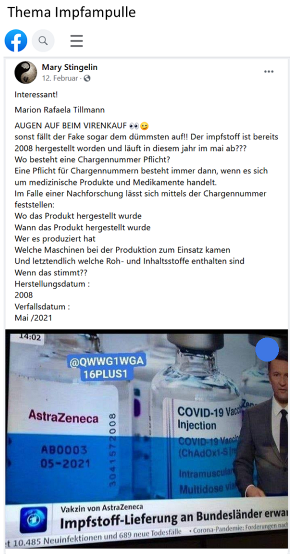
- I went to university.
- I like to travel to Croatia.
- I do not have any kids.
- I like reading and I read a lot.
- I have a driving license.
- **I consume dietary supplements.**

B.1.3 Fakes, facts, and fact-checks

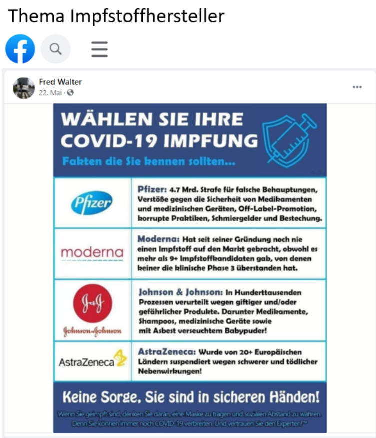
Translations of the figures can be found in Appendix B.1.5.

(a) Fake on Corona vaccines, Wave I

(b) Fake on Corona vaccines, Wave I



(c) Fake on Corona vaccines, Wave II



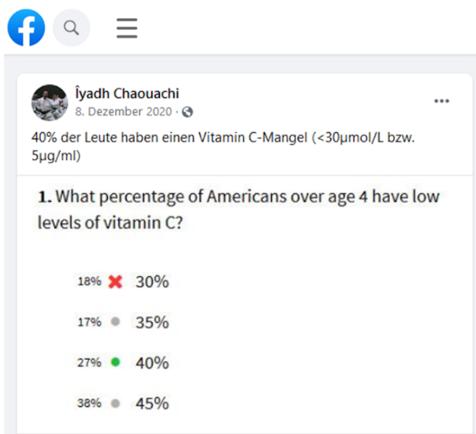
(d) Fake on Corona vaccines, Wave II

Figure B1: Fakes on Corona vaccines.

Thema Proteinbedarf

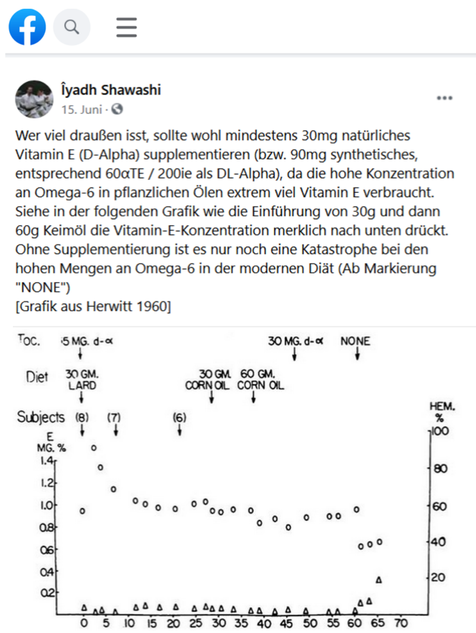
(a) Fake on Nutrition, Wave I

Thema Vitamin C



(b) Fake on Nutrition, Wave I

Thema Vitamin E



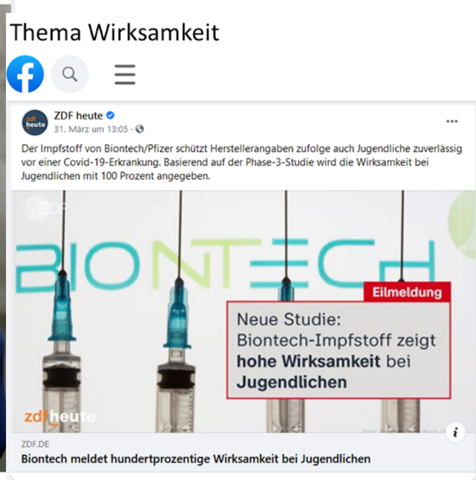
Thema Vitamin B12

(c) Fake on Nutrition, Wave II

(d) Fake on Nutrition, Wave II

Figure B2: Fakes on nutrition.

(a) Fact on Corona vaccines, Wave I



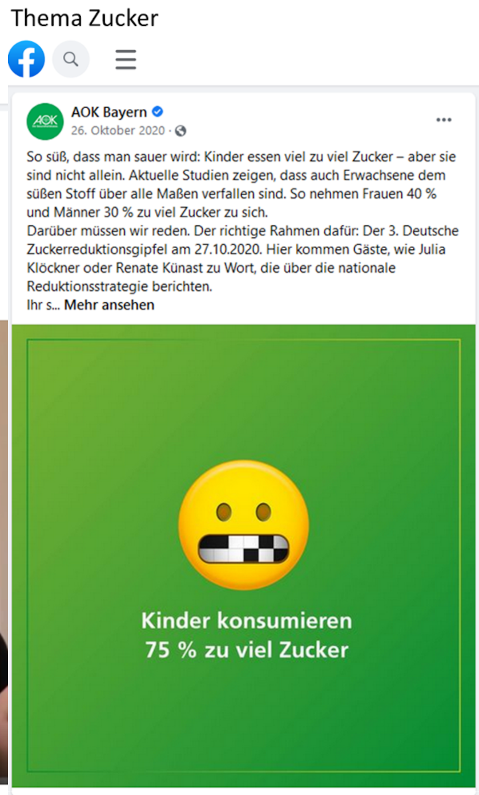
(b) Fact on Corona vaccines, Wave I

(c) Fact on Corona vaccines, Wave II

(d) Fact on Corona vaccines, Wave II

Figure B3: Facts on Corona vaccines.

(a) Fact on Nutrition, Wave I



(b) Fact on Nutrition, Wave I

(c) Fact on Nutrition, Wave II

(d) Fact on Nutrition, Wave II

Figure B4: Facts on Nutrition.

Thema Sana-Kliniken

Fakten-Check

CORRECTIV Recherchen für die Gesellschaft

Gerücht über Todesfälle nach Corona-Impfungen im Sana-Klinikum ist frei erfunden

Vor allem auf Facebook und Telegram kursiert aktuell das Gerücht, dass 50 Mitarbeitende im Berliner Sana-Klinikum zeitlich nach Covid-19-Impfungen gestorben seien. Für die Behauptungen gibt es keinerlei Belege. Das Klinikum weist sie als Falschmeldung zurück, das Landgericht Hamburg hat eine einstweilige Verfügung gegen die Verbreiterin des Gerüchts erlassen.

(a) Fact-check on Corona vaccines, Wave I

Thema Radioaktivität

Fakten-Check

AFP Faktencheck

Diese Geschichte über radioaktive Corona-Impfstoffe ist frei erfunden

Mehrere Hundert Facebook-User haben seit Mitte Dezember die Behauptung einer Person geteilt, die sich als Angestellte des Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM) ausgibt. Diese will eine Ampulle eines Impfstoffes aus dem Institut gestohlen und diese zu Hause untersucht haben. Ergebnis: Der Impfstoff soll angeblich leicht radioaktiv sein. Das BfArM lagert aber keinen Impfstoff, den Angestellte stehlen könnten. Das Institut ist nicht für Impfstoff-Zulassungen zuständig.

(b) Fact-check on Corona vaccines, Wave I

Thema Impfstoffhersteller

Fakten-Check

dpa • factchecking

Ungenaue Behauptungen über Impfstoffhersteller

23/05/2021, 02:51 PM (CEST)

Die Impfstoffe gegen das Coronavirus Sars-CoV-2 werden seit rund einem halben Jahr in vielen Ländern der Welt eingesetzt. Eine Übersicht auf Facebook stellt nun einige Behauptungen über die Hersteller der bislang in der EU zugelassenen Impfstoffe auf (hier archiviert). Was ist dran an den Warnungen?

Bewertung

Einige der Aussagen sind ungenau oder aber es fehlt wichtiger Kontext. Zum Beispiel wird in den meisten europäischen Ländern mit dem Corona-Impfstoff von Astrazeneca geimpft.

Fakten

Über den Impfstoff des Unternehmens **Astrazeneca** heißt es, dass er von mehr als 20 europäischen Ländern wegen schwerer und tödlicher Nebenwirkungen «suspendiert» worden sei. Richtig ist, dass viele Länder den Einsatz des Impfstoffs im März vorübergehend ausgesetzt hatten oder haben. Allerdings verzichten nur Dänemark und Norwegen dauerhaft auf den Impfstoff. Die meisten anderen europäischen Staaten setzen ihn inzwischen wieder ein, wenn auch meist eingeschränkt. In Deutschland und Italien etwa wird der Impfstoff von den zuständigen Stellen für Menschen über 60 empfohlen.

Über die Firma **Johnson & Johnson** wird in dem Instagram-Beitrag behauptet, dass sie «in Hunderttausenden Prozessen verurteilt» worden sei. Für eine derart hohe Zahl gibt es keine Belege. Richtig ist, dass Johnson & Johnson unter anderem wegen eines Skandals um Asbest in Babypuder mit vielen Klagen konfrontiert ist.

Dem Unternehmen **Moderna** wird unterstellt, dass keiner seiner Impfstoffe in Zulassungsstudien die sogenannte Phase 3 überstanden habe. Durch die Formulierung in der Vergangenheitsform wird ein Eindruck erweckt, als ob die Studien abgebrochen worden seien oder die Behörden eine Zulassung verweigert hätten. Dem ist jedoch nicht so: Ein Blick in eine offizielle Übersicht der US-Behörden zeigt, dass Studien zu anderen Moderna-Impfstoffen derzeit zum Teil noch laufen und sich noch in anderen Phasen befinden. Die Studien laufen hier lediglich deutlich länger als beim Corona-Impfstoff des Unternehmens.

Dessen Zulassung wurde zwar beschleunigt, in Europa etwa durch das sogenannte Rolling-Review-Verfahren. Aber auch dieser Impfstoff wurde vor der Zulassung mehr als 15 000 Testpersonen gespritzt, eine Größenordnung, die der vieler anderer Phase-3-Studien entspricht. Dabei wurde eine sehr hohe Wirksamkeit gegen Sars-CoV-2 festgestellt, die die Risiken deutlich übersteigt.

(c) Fact-check on Corona vaccines, Wave II

Thema Impfampulle

Fakten-Check

dpa • factchecking

Zahl auf Impfampulle fehlinterpretiert: kein Herstellungsdatum

03/03/2021, 06:26 PM (CET)

Eine Zahlenkombination auf einer Impfdosis soll angeblich beweisen, dass der Impfstoff schon lange vor Beginn der Corona-Pandemie entwickelt worden sein soll. Konkret geht es um ein Foto eines Fläschchens, das den Impfstoff der Firma Astrazeneca enthält (hier archiviert). Dieses Foto verbreitet sich auf Facebook. Auf dem Etikett ist der Name «AstraZeneca» zu sehen, außerdem «COVID-19 Vacci» und drei Zahlenkombinationen: «AB0003», «05-2021» und senkrecht an der Seite des Etiketts die Nummer «3041572008». Letztere wird in dem zum Foto gehörigen Beitrag als Chargennummer interpretiert. Angeblich soll sie verraten, wo und wann das Produkt hergestellt worden sei, wer es produziert habe sowie welche Maschinen und Inhaltsstoffe dabei zum Einsatz gekommen seien. Im Beitrag zu dem Foto heißt es zur Interpretation des Etiketts: «Herstellungsdatum : 2008» und «Verfallsdatum : Mai /2021».

Die seitlich stehende Nummer «3041572008» ist also nicht die Chargennummer. Auf dpa-Anfrage teilte Astrazeneca mit, dass es sich dabei um einen Komponenten-Code handelt, der für die jeweilige Produktionsstätte feststehend sei und für die Rückverfolgung in internen Systemen verwendet werde. Auf einen früheren Produktionsbeginn des Impfstoffs deute die Zahl nicht hin. «Die kommerzielle Produktion des Covid-19-Impfstoffs von Astrazeneca fing nicht vor Herbst 2020 an», teilte eine Sprecherin mit.

(d) Fact-check on Corona vaccines, Wave II

Figure B5: Fact-checks on Corona vaccines.

Thema Proteinbedarf

Fakten-Check

Deutsche Gesellschaft für Ernährung e.V.
Der Wissenschaft verpflichtet – Ihr Partner für Essen und Trinken

14. Gibt es gesonderte Empfehlungen von der DGE für die Proteinzufuhr für Sportler?

Für erwachsene Breitensportler*innen (4–5 Mal je Woche 30 Minuten körperliche Aktivität bei mittlerer Intensität) gibt es keine gesonderte Empfehlung. Zur Sicherstellung der Proteinversorgung reicht eine Zufuhr in Höhe der empfohlenen Zufuhr von 0,8 g Protein/kg Körpergewicht pro Tag aus.

Im ambitionierten Breitensport und Leistungssport (mind. 5 Stunden Training pro Woche) kann eine sportart- und belastungsspezifisch angepasste Proteinzufuhr den Trainingsprozess sinnvoll unterstützen und die Leistungsbereitschaft fördern. Über Zufuhrmenge, Art der Proteinquelle, optimale Aminosäurezusammensetzung sowie Zeitpunkt der Zufuhr wird teilweise kontrovers diskutiert. Die *International Society of Sports Nutrition* und das *American College of Sports Medicine* empfehlen je nach Sportart und Trainingsziel, -intensität, -umfang oder Wettkampfphase eine flexibel angepasste Proteinversorgung mit ca. 1,2–2,0 g/kg Körpergewicht pro Tag. Die Proteine sollten über den Tag verteilt und im Rahmen von Mahlzeiten und nicht als Supplemente zugeführt werden.

Im Positionspapier der Arbeitsgruppe Sporternährung der DGE zu Proteinzufuhr im Sport werden aktuelle Erkenntnisse zu physiologischen Wirkungen der Proteinzufuhr im Sport, unter besonderer Berücksichtigung des Zufuhrmenge- bzw. des Dosis-Wirkung-Aspekts, dargestellt.

(a) Fact-check on Nutrition, Wave I

Thema Vitamin C

Fakten-Check

Deutsche Gesellschaft für Ernährung e.V.
Der Wissenschaft verpflichtet – Ihr Partner für Essen und Trinken

3. Gibt es hierzulande einen Vitamin-C-Mangel?

In industrialisierten Ländern kommen Vitamin-C-Mangelzustände praktisch nicht mehr vor. Klassische klinische Vitamin-C-Mangelzustände sind beim Säugling die Moeller-Barlow- Krankheit und beim Erwachsenen der Skorbut (früher oft als „Seefahrerkrankheit" beschrieben). Dabei sind die Knochenbildung und das Wachstum beim Säugling und Kind gestört. In späteren Lebensabschnitten sind die Symptome schlechte Wundheilung, Gelenkschmerzen, Infektionen, Neigung zu Blutungen in der Haut, den Schleimhäuten, der Muskulatur und den inneren Organen sowie Zahnausfall. Diese Störungen treten bei Erwachsenen nur bei dauerhaft fehlender Vitamin-C-Zufuhr auf. Bereits 10 mg Vitamin C pro Tag verhindern Skorbut.

(b) Fact-check on Nutrition, Wave I

Thema Vitamin B12

Fakten-Check

verbraucherzentrale

Was steckt hinter der Werbung zu Vitamin B12?

Immer wieder hört man von einem weit verbreiteten Vitamin B12-Mangel in der Bevölkerung. Wissenschaftliche Belege für diese Behauptungen gibt es keine. Sowohl Berechnungen zur Vitamin B12-Versorgung als auch Untersuchungen im Blut zeigten, dass ein Vitamin B12-Mangel selten ist. In der Regel führt die westliche Ernährungsweise mit einem hohen Anteil an tierischen Lebensmitteln eher zu einer Vitamin B12-Überversorgung.

(c) Fact-check on Nutrition, Wave II

Thema Vitamin E

Fakten-Check

GESUNDHEIT.GV.AT
Öffentliches Gesundheitsportal Österreichs

Wie viel Vitamin E brauchen wir?

Der genaue Bedarfswert ist nicht bekannt. Der Schätzwert für eine angemessene Zufuhr für Erwachsene (25 bis < 51 Jahre) liegt pro Tag nach den D-A-CH-Referenzwerten bei 12 mg (Frauen) und 14 mg Vitamin E (Tocopherol-Äquivalente). Für **schwangere Frauen** liegt der Schätzwert für eine angemessene Zufuhr bei 13 mg und für **Stillende** bei 17 mg pro Tag.

Näheres zu allen Altersklassen bzw. Personengruppen sowie Geschlecht erfahren Sie in den D-A-CH-Referenzwerten. Weitere Informationen erhalten Sie unter Deckung des Tagesbedarfs an Vitamine.

Zu viel/zu wenig Vitamin E?

- **Ein Zuviel** an Vitamin E ist selten, dennoch sollten hohe Dosierungen (Supplemente) über einen längeren Zeitraum vermieden werden, da sie u.a. Magen-Darm-Probleme und ein erhöhtes Blutungsrisiko verursachen können. Als tolerierbare Gesamtzufuhr gelten 300 mg pro Tag (Tolerable Upper Intake Level / EFSA).

(d) Fact-check on Nutrition, Wave II

Figure B6: Fact-checks on Nutrition.

B.1.4 (Un-)trustworthy elements in fakes and facts

Translations of the figures can be found in Appendix B.1.5.

Figure B7: Exemplary types of elements that participants could indicate as "trustworthy" or "not trustworthy"

B.1.5 Translations of the images

This section provides translations for the images in Sections B.1.3 and B.1.4.

B.1.5.1 Translation of Figure B.1

Translation of Figure B.1 (a) Topic on sana-clinics

Heike Tamm:

"So much on that. The nursing staff is becoming scarce, it's dying."

Helena Wolf:

Sana-clinics

** Who wants a job there? **

The chances are better than ever.

Reason:

Briefly after "vaccination" of staff members, 50 people have died by now.

The clinics management forced everyone to remain silent.

Risking loosing ones job, if one was heathy enough to survive the "vaccination".

Source: No newspaper. Now news station! But: A Sana clinic Berlin-staff member-family member is a colleage of my husbands in dementia-care.

Translation of Figure B.1 (b) Topic on radioactivity

Christian I_forgot:

I work at BfArM (federal institute for drugs and medical products) and I had the opportunity to smuggle an ampoule of the vaccine secretly out of the building to analyse it at home. I discovered that this alleged vaccine glows in the dark and is slightly radioactive.

At home I measured approximately 70 microsievert per hour. This is enough to cause diseases as well as genetic damages.

There might be radium in this vaccine.

Alert everyone about this madness !!!

Here is a picture of the "vaccine" that I pulled from the ampoule with an injection needle:

(picture)

Translation of Figure B.1 (c) Topic on vaccine ampoules

Mary Stingelin:

Interesting!

Marion Rafaela Tillmann:

OPEN YOUR EYES WHEN BUYING VIRUSES (eye emoji, winking emoji)

Otherwise even the dumbest would see the fake!! The vaccine was already produced in 2008 and expires this year in May???

Where does a batch number duty exist?

An obligation to a batch number exists, whenever dealing with medical devices or drugs.

In case of an inquiry, the batch number provides:

Where the product was produced

When the product was produced

Who produced it

Which machines were used for production

And finally, which raw materials and ingredients it contains

If this is true??

Production date: 2008

Expiration date: May/2021

(in the picture:

Vakzin from AstraZeneca

Vaccine-delivery to states expe[cted]

10.485 new infections and 689 new deaths. Corona-pandemic: Demands [...])

Translation of Figure B.1 (d) Topic on producers of vaccine

Fred Walter:

(in the picture:

CHOOSE YOUR COVID-! VACCINE)

Facts you all should know about...

Pfizer: 4.7 billion penalty for false allegations, violation of the safety of drugs and medical devices, off-label-promotion, corrupt practices, bribery and bribe money.

Moderna: Hasn‘t brought a vaccine onto the market since its foundation, even though there were more than nine candidates. None of those survived the clinical phase 3.

Johnson & Johnson: Convicted in hundred thousand processes because of toxic and/or dangerous products. Including drugs, shampoos, medical devices as well as baby powder polluted with asbestos!

AstraZeneca: Was suspended by 20+ European countries because of severe and deadly side effects!

No worries, you are in save hands!

Once you are vaccinated, remember social distancing and to wear a mask.

As you can still spread COVID-19. And trust the experts!

B.1.5.2 Translation of Figure B.2

Translation of Figure B.2 (a) Topic on protein demand

Clever fit Ludwigsfelde is here: clever fit Ludwigsfelde

Do you achieve your daily demand of protein?

Protein is by far the most important macronutrient, because how are new muscles supposed to grow, if there is no construction material?

The general advice is: a daily dose of 2g per kg body weight

Do you achieve to meet your daily protein demand?

(Hashtags)

Translation of Figure B.2 (b) Topic on Vitamin C

Îyadh Chaouachi:

(picture already in English)

Translation of Figure B.2 (c) Topic on Vitamin B12

Dr. Rath health-foundation:

Up to 40 percent of adults show a shortage of vitamin B12, which has various consequences.

ISSUU.COM

DR. RATH HEALTH LETTER – version 09/21 – June 2021 – VITAMIN B12: Shortage can cause tinnitus

Up to 40 Percent of adults show a shortage of vitamin B12, which has various consequences. If there is a lack of Vitamin B12, the vessel and the nervous system of the ear can be damaged for example and a tinnitus might be caused. The symptoms are diffuse and the course of disease often insidious – d...

Translation of Figure B.2 (d) Topic on Vitamin E

Îyadh Shawashi:

Those, who often eat outside, should probably supplement at least 30 mg natural vitamin E (D-Alpha) (resp. 90mg synthetic, equalling 60 aTE/200ie as DL-alpha), as the high concentration of Omega-6 in plant based oils consumes extreme amounts of vitamin E.

See how in the following graphic the introduction of 30g and then 60g seed oil presses down the vitamin E concentration notably.

Without supplementation it is just a catastrophe with the high amounts of Omega-6 in modern diets

(From mark "NONE")

[Graphic from Herwitt 1960]

B.1.5.3 Translation of Figure B.3

Translation of Figure B.3 (a) Topic on Sputnik V

ZDF today:

Even before a possible permit in the EU, Bavaria secured more than two million doses of the Russian vaccine Sputnik V for itself: (Link)

(In the picture: Bavaria secures 2.5 million doses of Sputnik V for itself)

Translation of Figure B.3 (b) Topic on Effectiveness

ZDF today:

The vaccine from Biontech/Pfizer protects according to manufacturer information adolescence responsibly against a Covid-19-infection, too. Based on the phase-3-study the effectiveness in adolescence is labelled with a hundred percent.

(in the picture: Breaking news
New study: Biontech-vaccine shows high effectiveness in adolescence)

ZDF.DE
Biontech announces a hundred percent effectiveness in adolescence.

Translation of Figure B.3 (c) Topic on Staying Away

FAZ.NET – Frankfurt general newspaper:

In Saarland more than 100 missed vaccination appointments of medical staff cause furore:

(Picture)

FAZ.NET| FROM FAZ.NET – Frankfurt general newspaper Why do more than 100 doctors not get vaccinated?

Translation of Figure B.3 (d) Topic on vaccination complications

Rheinische Post:

Reasons seem to be complications with vaccination.

(Picture)

RP-ONLINE.DE

18 vaccinees want to take the state NRW into account
Because of complications, 18 vaccinees currently want to take the state NRW into accou[nt]

B.1.5.4 Translation of Figure B.4

Translation of Figure B.4 (a) Topic on plastic

AOK Bavaria:

Life in plastic … isn’t fantastic! (littering emoji)
Did you test once how much of an example you are, too, and inspected your apartment on accruing plastic?
You will probably be as shocked as Inka. (shocked emoji)
Editor‘s note: On average we consume approximately 5g micro plastic per week.
You can find tipps on zero waste shopping here: (link)

(In the picture:Plastic is everywhere
Click to zoom
#ItStartsWithYou)

Translation of Figure B.4 (b) Topic on sugar

AOK Bavaria:

So sweet, that one could turn sour: Children are eating way too much sugar — but they are not alone. Current studies show, that adults, too, have fallen excessively for the sweet stuff.
Due to that 40% of women and 30% of men are eating too much sugar.
We need to talk about that. The right frame for it: The 3rd German sugar-reduction-summit on 27.10.2020. Here, guests like Julia Klöckner or Renate Künast will report on the national reduction strategy. Their … Load more

(In the picture: Children consume 75% too much sugar)

Translation of Figure B.4 (c) Topic on moving

Federal institute for nutrition:

80 percent of adolescents are moving less than recommended by the world health organisation!

The network "Healthy into life — network young family" as well as IN SHAPE — Germanys initiative for healthy nutrition and more movement provide various information, recommendations, material and recipes on their websites

(picture)

BZFE.DE
Less Couch-Potatos –BZfE
Federal institute for nutrition (BZfE): The competence- and commu...

Translation of Figure B.4 (d) Topic on nutrition scores

Upfit:

Time for a food comparison (boxing glove emoji):
Do you regularly eat chia or linseeds? And if so, with what do you like them best? (smiley emoji)
Chia seeds are the Latin American pendant to the European linseeds. The nutrition scores of both of them don't vary much. Linseeds have 471 kcal for every 100g, chia seeds 475 kcal.
Chia and linseeds deliver high-value protein and omega-3- and omega-6-fatty acids. (strong arm emoji)
The especially high share of calcium and vitamin A in the chia seeds sets those apart. Linseeds meanwhile contain a lot of potassium and iron and zinc are contained as well.
All in all, both seeds provide important nutritions and are very healthy. Those who prefer a regional origin of their food should choose the linseeds. These are also cheaper than the chia seeds most of the time.
#projectlivehealthy #fatlogicfree #eatingbetter #looseweightgermany #dietary-diary #easycooking #eatinghealthy #thoughtfuleating #looseweightwithfun #youarewhaty-oueat #looseweightintuitively #lowincaloriesrecipes

(In the picture: food comparison
Chia seeds | line seeds
Learn more on:
UPFIT)

B.1.5.5 Translation of Figure B.5

Translation of Figure B.5 (a) Topic on Sana-clinics

Fact-check

CORRECTIV
Research for the society

Rumor about mortality after Corona-vaccination in Sana-clinic is fictitious.

Especially on Facebook and Telegram there currently is a rumor about the death of 50 staff members of Berlin‘s Sana-clinic shortly after Covid-19-vaccinations. There is no evidence for these claims. The clinic dismisses them as false reports, the state court of Hamburg has issued an injunction against the propagator of this rumor.

Translation of Figure B.5 (b) Topic on radioactivity

Fact-check

AFP fact-check

This story about radioactive Corona-vaccinations is fictitious.

Several hundred Facebook-User shared the claim of a person, who impersonated a staff member of the federal institute for drugs and medical devices (BfArM), since mid-December. This person claims to have stolen an ampoule of the vaccine from the institute and to have tested it at home.

Result: The vaccine is supposedly slightly radioactive. The BfArM however isn’t storing any vaccines staff members could steal. The institute isn’t responsible for vaccine-permits.

Translation of Figure B.5 (c) Topic on vaccine manufacturer

Fact-check

dpa factchecking

Uncertain claim about vaccine manufacturer

The vaccines against the Coronavirus Sars-CoV-2 are being used for about six months in many countries around the world. An overview on Facebook now raises some claims about the manufacturers of the vaccines that are authorized in the EU up to now (archived here). What‘s up with the warnings?

Evaluation

Some of the statements are vague or there is important context missing. For example most European countries use the vaccine from Astrazeneca to vaccinate.

Facts

About the vaccine of the company Astrazeneca is said that it is suspended by more than 20 European countries because of severe or deadly side effects. True is, that many countries paused or have paused the use of the vaccine temporarily in March. However, only Denmark and Norway renounced the vaccine permanently. Most of the other European states are already using the vaccine again, even though mostly with some restrictions. In Germany and Italy the vaccine is additionally only recommended to people over 60 by the responsible bodies.

About the company Johnson & Johnson the Instagram-post states, that it "was convicted in hundred thousand processes". There is no proof for a number this high. It is true, that Johnson & Johnson is among other things confronted with many complaints because of a scandal about asbestos in baby powder.

It is assumed that none of the vaccines from the company Moderna surpassed the so-called phase 3 of their permission-studies. Due to the formulation in past tense the impression arises that the study was cancelled or that the authorities denied the permit. It is not like that though: A glance into an official overview by the US-authorities shows, that studies on other Moderna-vaccinations are currently still running or are still in other phases. The studies are just running a lot longer here than the companies Corona-vaccine.

This permit was indeed accelerated, in Europe for example by the rolling-review-procedure. Nevertheless more than 15 000 test subjects, a size similar to many other phase-3-studies, were injected with this vaccine before permission. With this a large effectiveness against Sars-CoV-2 was determined, which exceeds the risks by far.

Translation of Figure B.5 (d) Topic on vaccine ampoule

Fact-check

dpa factchecking

Number on vaccination ampoule misinterpreted: no production date

A combination of numbers on a vaccination dose supposedly proves that the vaccine has been developed long before the Corona-pandemic. In detail this is about a picture of a small flask that contains the vaccine of the company Astrazeneca (archived here). This picture spread on Facebook. On the label the name "AstraZeneca" is depicted, additionally "COVID-19 Vacci" and three combination of numbers: "AB0003", "05-2021" and vertically on the side of the label the number "3041572008". The last is interpreted as batch number in the post concerning the picture.

It is supposed to reveal, where and when the product was produced, who produced it and as which machines and ingredients have been used. In the post with the picture it was said: “production date: 2008” and “expiration date: May /2021”

The vertical number “3041572008” on the side is not the batch number. In response to dpa‘s request, Astrazeneca replied that the number series is a component-code, which is set for each production site and is needed in internal systems for trace back. The number is not referring to an earlier start of production. “the start of commercial production of the Covid-19-vaccine from Astrazeneca was not before fall 2020”, a speaker announced.

B.1.5.6 Translation of Figure B.6

Translation of Figure B.6 (a) Topic on protein demand

Fact-check

German society of nutrition e.V. (e. V. = enlisted association)

Committed to science – your partner for food and drinks

14. Are the special recommendations about protein intake for athletes by the German society of nutrition (DGE)?

For grown up general athletes (30 minutes of physical activity with medium intensity 4-5 times a week) there are no special recommendations. To ensure meeting ones protein demands, the recommended amount of 0.8 g protein per kg body weight per day is enough.

In ambitious mass sport and competitive sports (minimum of 5 hours training per week) a protein intake that is adjusted specifically according to the kind and load of sport, might support the process of training and promote the willingness to perform. There is a partly controversial debate about the amount of intake, type of protein source, optimal aminoacid composure as well as time of intake. The International Society of Sports nutrition and the American College of Sports Medicine recommend flexibly adjusted protein supply based on type of sport and goal, intensity and load of training with approximately 1.2 – 2.0 g/kg body weight per day. The proteins should be evenly distributed over the day and be consummated during meals and not as supplements.

In the policy paper of the working group on sports nutrition of the DGE about protein intake in sports current findings about psychological effects of protein intake in sports, with special consideration of the amount of intake- and dose-effect-aspects, are being discussed.

Translation of Figure B.6 (b) Topic on Vitamin C

Fact-check

German society of nutrition e.V. (e. V. = enlisted association)
Committed to science – your partner for food and drinks

3. Is there a vitamin-C-lack in this country?

In industrialised nations vitamin-C-lacks barely exist anymore.

Common clinical vitamin-C-lacks are the Moeller-Barlow disease in infants and the Skorbut (also known as the "sailor‘s disease") in adults. With these diseases the formation of new bones and growth in general is disturbed in infants and children. Later on the symptoms are bad healing of wounds, joint pain, infections, tendency to bleeding within the skin, the mucosae, the muscles and inner orangs, as well as loosing ones teeth. In adults the disorders only appear if there is a continuous lack in vitamin-C supply. 10 mg vitamin C a day can already stop Skorbut.

Translation of Figure B.6 (c) Topic on Vitamin B12

Fact-check

User headquarters

What‘s behind the advertisement for vitamin B12?

Regularly one hears about a widespread vitamin B12-lack in general population. There is no scientific proof for this claim. Calculations about the vitamin B12 supply as well as examinations of blood showed, that a lack of vitamin B12 is rare. Usually the western diet with a large portion of animal products rather causes a vitamin B12 overflow.

Translation of Figure B.6 (d) Topic on Vitamin E

Fact-check

health.GV.AT
open health portal Austria

How much vitamin E do we need?

The exact amount is unknown. An estimate for adults (25 till $<$ 51 years) is according to D-A-CH-reference values 12 mg of vitamin E (Tocopherol-equivalents) a day for women and 14 mg for men. For pregnant women an appropriate estimate is 13 mg and 17 mg a day for nursing women.

You can find more about all age-levels and groups of people as well as genders in the D-A-CH-reference values. More information under achieving the daily demand of vitamins.

Too much/not enough vitamin E?

A too much vitamin E is rare, nevertheless high doses (supplements) over a longer period of time should be avoided, as they might cause gastrointestinal problems and a heightened risk of bleeding. About 300mg per day are considered a tolerable upper intake level.

B.1.5.7 Translation of Figure B.7

Figure B.7

Helena Wolf

Sana-clinics

** Who wants a job there? **

The chances are better than ever.

Reason: Briefly after “vaccination” of staff members, 50 people have died by now.

The clinics management forced everyone to remain silent. Risking loosing ones job, if one was heathy enough to survive the “vaccination”.

Source: No newspaper. Now news station! But: A Sana clinic Berlin-staff member-family member is a colleage of my husbands in dementia-care.

B.2 Pre-analysis plan

B.2.1 Main experiment

We registered the following pre-analysis plan in the AEA Registry under registry number AEARCTR-0008199. Please see https://doi.org/10.1257/rct.8199-1.0 for further details.

Important note: We initially referred to the "short- and longer-term effects" of our interventions, but changed the wording to "immediate and short-term effects". Hence, when the pre-analysis plan refers to the "short-run effects", it refer to what we call "immediate effects" in the current version of the paper. Moreover, when the pre-analysis plan refers to the "longer-run effects", it refers to what we call "short-run effects" in the current version of the paper.

Interventions

1. Fact-check intervention

2. Media literacy intervention

Intervention Start Date

2021-09-10

Intervention End Date

2021-10-15

Primary Outcomes (end points)

1. Participants' factual knowledge on topics covered by the "fake news".

2. Perceived credibility of the "fake news".

3. Behavioral intentions:

    (a) Probability to get Covid-19-vaccine. Note: participants who are not vaccinated yet will be asked about their probability to get vaccinated as soon as they re-ceive an offer. Participants who are already vaccinated will be asked about the probability to get a booster in case the gov-ernment recommended this.

(b) Probability to consume dietary supplements

Secondary Outcomes (end points)

1. Participants' factual knowledge on topics covered by the facts.
2. Perceived credibility of facts.
3. Time spent with the "Tips how to spot fake news".
4. Time spent with the fact-checks.

Experimental design

We will expose the participants to fakes and facts on Corona vaccines and on nutrition that we retrieve from Facebook.

The participants will be randomly divided into five groups:

- Participants who are not exposed to any fakes or facts at all.
- Participants who are exposed to fakes and facts.
- Participants who are exposed to fakes and facts and fact-checks that debunk the fake news on vaccines (not the fakes on nutrition) → Fact-check intervention.
- Participants who are shown Facebook's "Tips how to spot false news" and are afterwards exposed to fakes and facts → Media literacy intervention.
- Participants who are exposed to facts and the fact-checks on vaccines.

To study the short-term effects of the fact-check and the media literacy intervention, we will compare participants' factual knowledge, perceived credibility of the "fake news", and their behavioral intentions w.r.t. Corona vaccines to participants who saw the fakes and facts, but did not receive an intervention.

To examine potential spillover effects of the fact-check and media literacy intervention, we show the participants fakes and facts on an unrelated topic (dietary supplements) without further interventions. Then, we will again compare factual knowledge, perceived credibility of the "fake news", and behavioral intentions of participants in the fact-check and in the media literacy group to participants who saw the fakes and facts, but did not receive an intervention.

Finally, to study the long-term effects of the fact-check and media literacy intervention, we re-invite our participants about one week later and repeat our analyses.

Estimation procedure Our baseline regressions will be OLS estimations with treatment group dummies as main explanatory variables. The dependent variables are perceived credibility of the "fake news", factual knowledge on Corona vaccines and dietary supplements, and self-reported probabilities to get vaccinated and to consume dietary supplements. Control variables include age, gender, family status, income, education, "big 5", political preferences, and prior knowledge on current events, health, and nutrition. As participants may skip over the fact-checking and media literacy intervention, our baseline analysis is equivalent to an ITT analysis. To obtain estimates for the LATE of our interventions, we will also conduct an IV estimation, where we use the time spent with the fact-checks or the "Tips how to spot false news" as endogenous explanatory variable and assignment to treatment as instrument.

Hypotheses We hypothesize the following effects:

1a In the short-run, the fact-checking intervention reduces the credibility of and increases factual knowledge about the corrected "fake news" as compared to participants without intervention.

1b In the short-run, participants who received the fact-checking intervention are more likely to state that they are willing to get vaccinated against Covid-19 than participants without intervention.

2a In the short- and in the long-run, the media literacy intervention reduces the credibility of and increases factual knowledge about all "fake news" as compared to participants without intervention.

2b In the short- and in the long-run, participants who received the media literacy intervention are more likely to state that they are willing to get vaccinated against Covid-19 and abstain from dietary supplements than participants without intervention.

2c The long-run effects of the media literacy intervention are smaller than its short-run effects.

The idea is as follows:

While fact-checking is likely to be an effective tool to debunk the corrected messages themselves, its effect may be limited to these messages. Thus, the intervention reduces the credibility of the "fake news" on Corona vaccines that we display in Session 1 as compared to participants without intervention. As a result, factual knowledge on Corona vaccines is larger than in the group without intervention and participants are more likely to get vaccinated. In contrast to that, we do not expect to see spillover effects on "fake news" about dietary supplements or on "fake news" on Corona vaccines in Session 2. The media literacy intervention conveys general skills that can be broadly applied. Hence, we expect that the intervention reduces the credibility of "fake news" on Corona vaccines and dietary supplements that we display in Session 1 and 2 as com-pared to participants without intervention. As a result, factual knowledge on Corona vaccines and dietary supplements is larger than in the group without intervention and participants are more likely to get vaccinated and less likely to consume supplements. As some participants may have forgotten the "Tips how to spot false news", we expect the effects to be smaller in Session 2 than in Session 1.

Subgroup analyses To date, about half of the German population is already fully vaccinated against Covid-19 and we expect a similar fraction of our participants to be vaccinated, too. As vulnerability to "fake news" may depend on prior knowledge and experiences, we will split the sample into vaccinated and non-vaccinated participants and explore heterogeneous effects. Similarly, we will partition our sample into participants who agree with the official Corona remedies vs. those who don't to see if and how prior beliefs and attitudes moderate the effects of our interventions. Finally, an interesting subgroup analysis will be to partition participants into those who spend much and those who spend few time on social media to see if experience and prior exposure to "fake news" from social media plays a role.

Randomization Method

The participants will be randomly allocated into the groups by the market research company that we are going to cooperate with.

Randomization Unit

Individual

Was the treatment clustered?

No

Sample size: planned number of clusters

NA

Sample size: planned number of observations

3000 participants (individuals) divided into five groups for the first wave. Planning with 20% attrition after 1 week: about 2400 participants (individuals) divided into five groups for the second wave.

Sample size (or number of clusters) by treatment arms

600 participants (individuals) per group for the first wave. Planning with 20% attrition after 1 week: about 480 participants (individuals) per group for the second wave.

B.2.2 Robustness checks

For our robustness checks, we registered the following pre-analysis plan in the AEA Registry under registry number AEARCTR-0013629. Please see `https://www.socialscienceregistry.org/trials/13629` for further details.

Interventions

1. Fact-check intervention
2. Media literacy intervention

Intervention Start Date

2024-07-08

Intervention End Date

2024-08-19

Primary Outcomes (end points)

Perceived credibility of the "fake news"

Secondary Outcomes (end points)

1. Perceived credibility of facts
2. Time spent with the "Tips how to spot fake news"
3. Time spent with the fact-checks

4. Time spent with the Facebook postings (fakes and facts)

5. Time spent with the credibility questions

Experimental design

We will expose the participants to fakes and facts on environmental and health-related topics that we retrieve from Facebook.

The participants will be randomly divided into six groups:

- Participants who are exposed to high-credibility fakes and facts → Control group high
- Participants who are exposed to low-credibility fakes and facts → Control group low
- Participants who are exposed to high-credibility fakes and facts and fact-checks that debunk the fake news → Fact-check intervention high
- Participants who are exposed to low-credibility fakes and facts and fact-checks that debunk the fake news → Fact-check intervention low
- Participants who are shown Facebook's "Tips how to spot false news" and are afterwards exposed to high-credibility fakes and facts → Media literacy intervention high
- Participants who are shown Facebook's "Tips how to spot false news" and are afterwards exposed to low-credibility fakes and facts → Media literacy intervention low

The environment-related fakes participants are exposed to have been rated as more-credible or less-credible in a pre-study. Participants in the "high" groups are exposed to two fakes which have been rated as more-credible, while participants in the "low" groups see two fakes which have been rated as less-credible. The facts and health-related fakes and facts are identical in all groups. The order in which the fakes and facts are displayed will be randomized. We also randomize whether participants in the fact-check group see the fact-check before or after the corresponding fake.

To study the effects of the fact-check and the media literacy intervention, we will compare participants' perceived credibility of the "fake news" to participants who saw the fakes and facts, but did not receive an intervention (always comparing the high groups to one another and the low groups to the other low groups). To study the effect of the degree of credibility we will compare the coefficients for the treatment effects in the high groups to the coefficients for the treatment effects in the low groups.

Estimation procedure Our main regressions estimate the effect of the treatment group dummies on the perceived credibility of "fake news". Control variables include age, gender, family status, income, education, "big 5", political preferences, and prior knowledge on current events, health, and nutrition.

Hypotheses We hypothesize the following effects:

- The fact-checking intervention reduces the credibility of the "fake news" as compared to participants without intervention.
- The media literacy intervention reduces the credibility of the "fake news" as compared to participants without intervention.
- The effects of both interventions are higher for fakes which are ex-ante rated as less credible.

Randomization Method

Computer.

Randomization Unit

Individual; within-individual for the analysis of the degree of credibility

Was the treatment clustered?

No

Sample size: planned number of clusters

NA

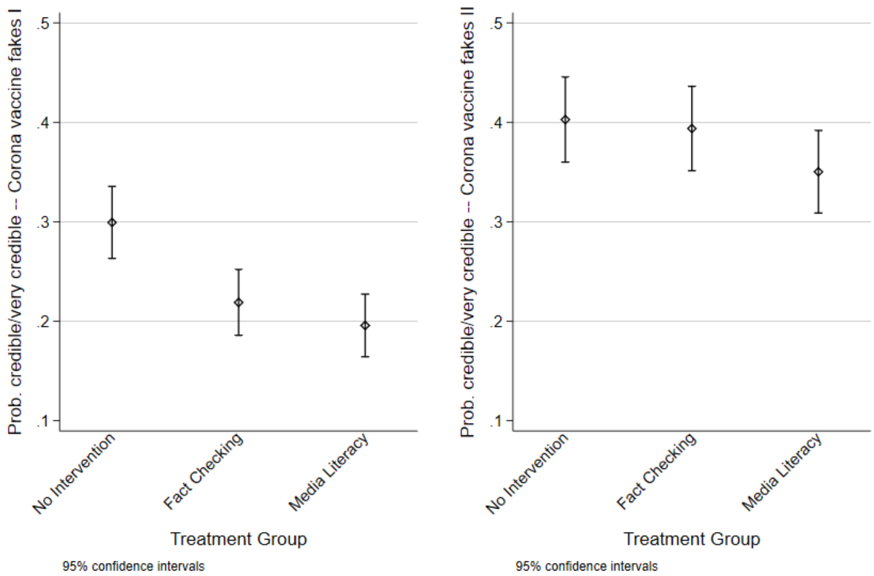
Sample size: planned number of observations

1800 participants (individuals)

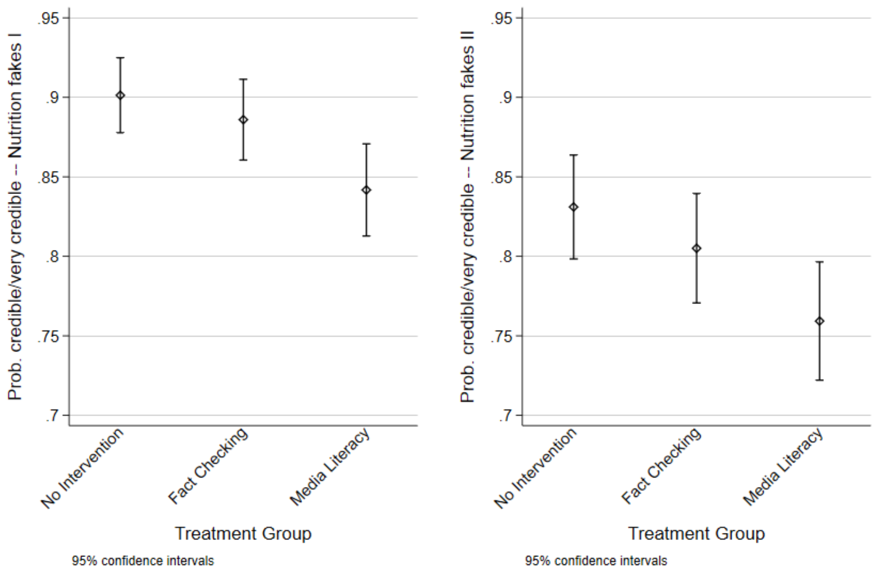
Sample size (or number of clusters) by treatment arms

600 high control group, 600 low control group, 600 high fact-check group, 600 low fact-check group, 600 high media literacy group, 600 low media literacy group (3600 total)

B.3 Omitted figures

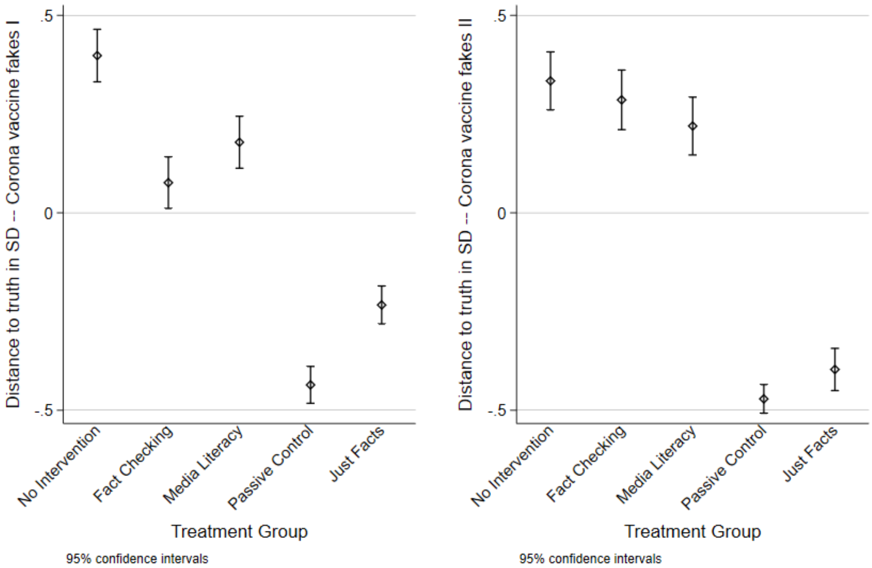


(a) Average credibility of **Corona fakes** per treatment group and survey wave

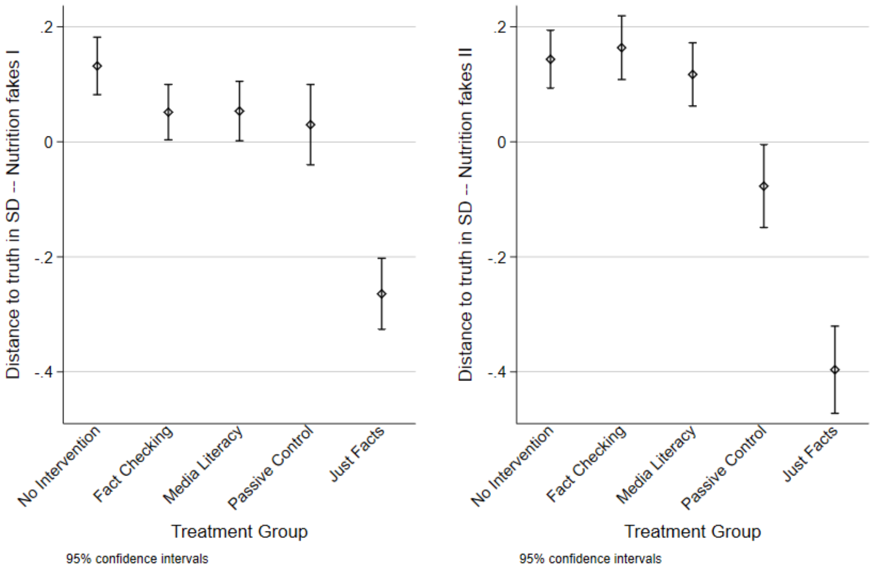


(b) Average credibility of **nutrition fakes** per treatment group and survey wave

Figure B8: Average credibility of fakes.

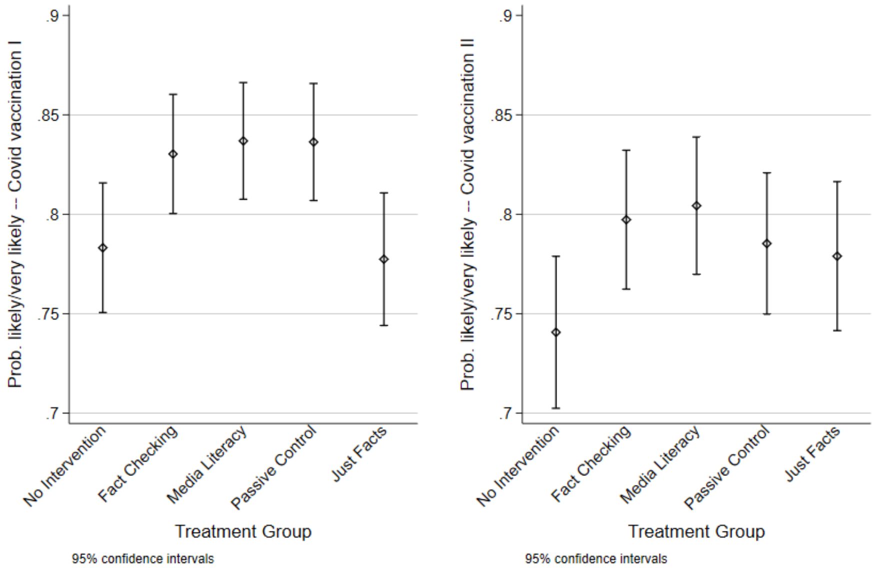


(a) Average standardized distance to the correct answers to questions on **Corona fakes** per treatment group and survey wave.

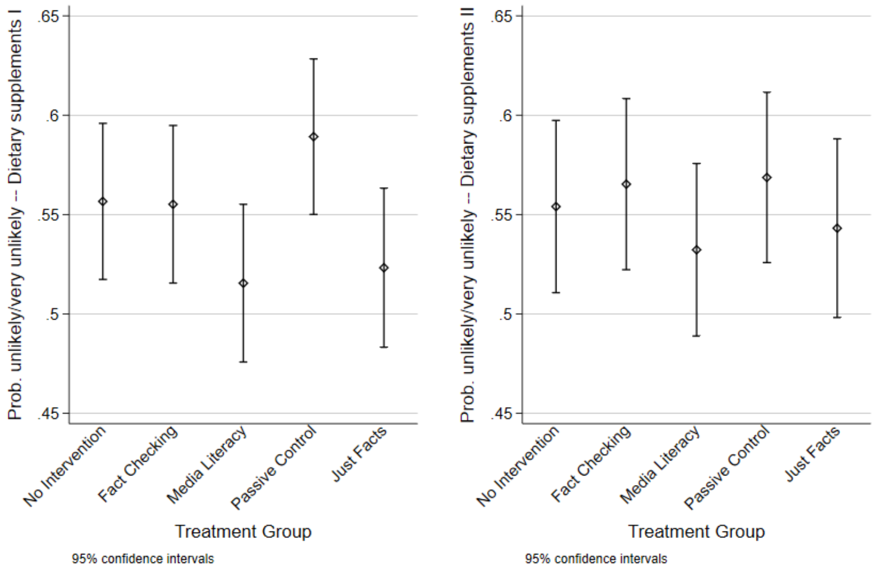


(b) Average standardized distance to the correct answers to questions on **nutrition fakes** per treatment group and survey wave.

Figure B9: Average standardized distance to the correct answers to the factual knowledge questions on fakes.

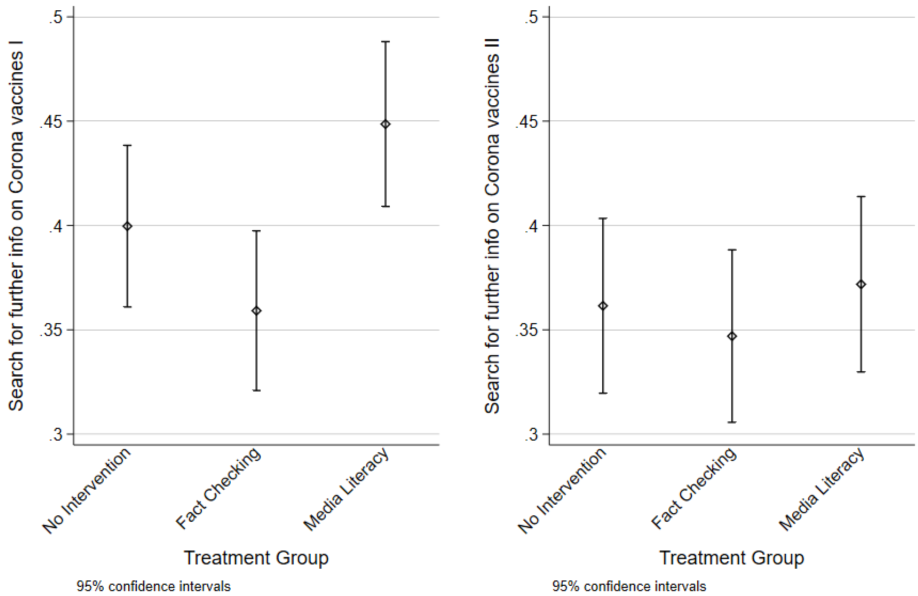


(a) Average reported probability to get vaccinated or boostered against Corona per treatment group and survey wave.

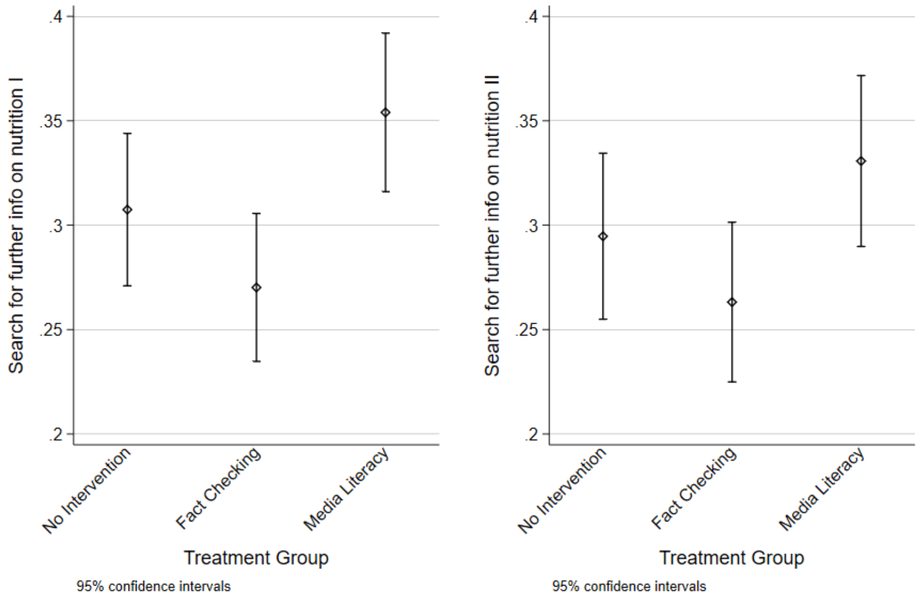


(b) Average reported probability to consume dietary supplements per treatment group and survey wave.

Figure B10: Average reported probability to get vaccinated or boostered against Covid-19 and to consume dietary supplements.

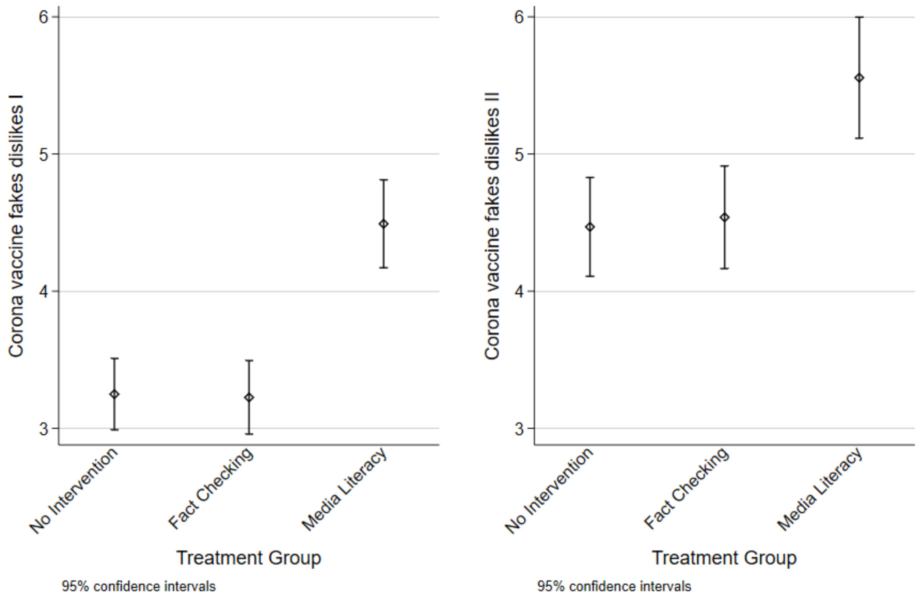


(a) Share of participants who searched for further information on Corona vaccines.

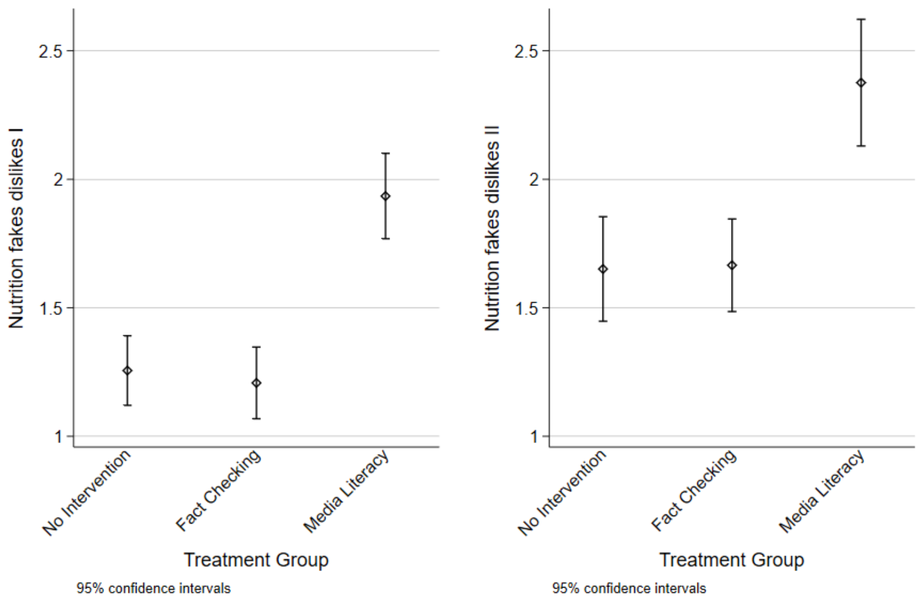


(b) Share of participants who searched for further information on nutrition.

Figure B11: Share of participants who searched for further information online.

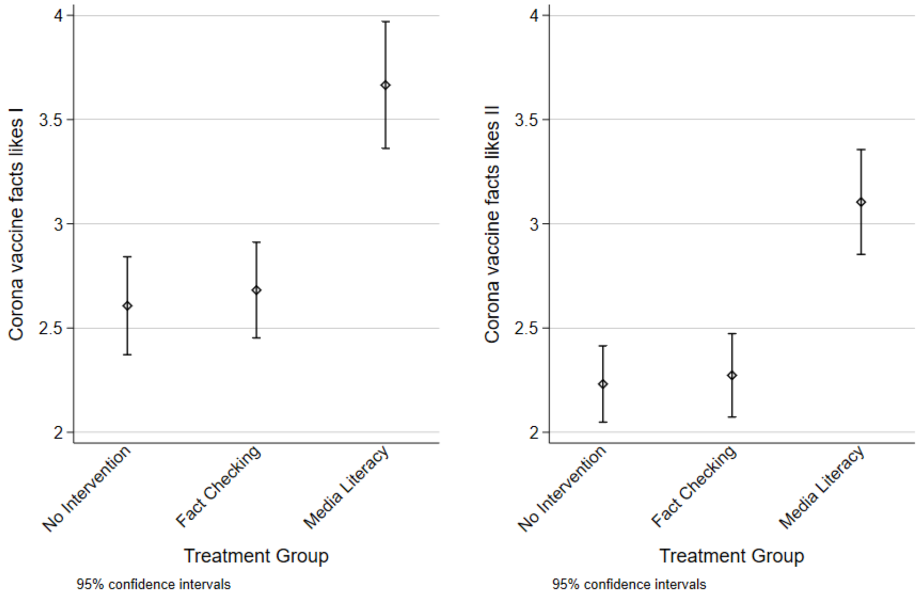


(a) Average number of untrustworthy elements for fakes on Corona vaccines.

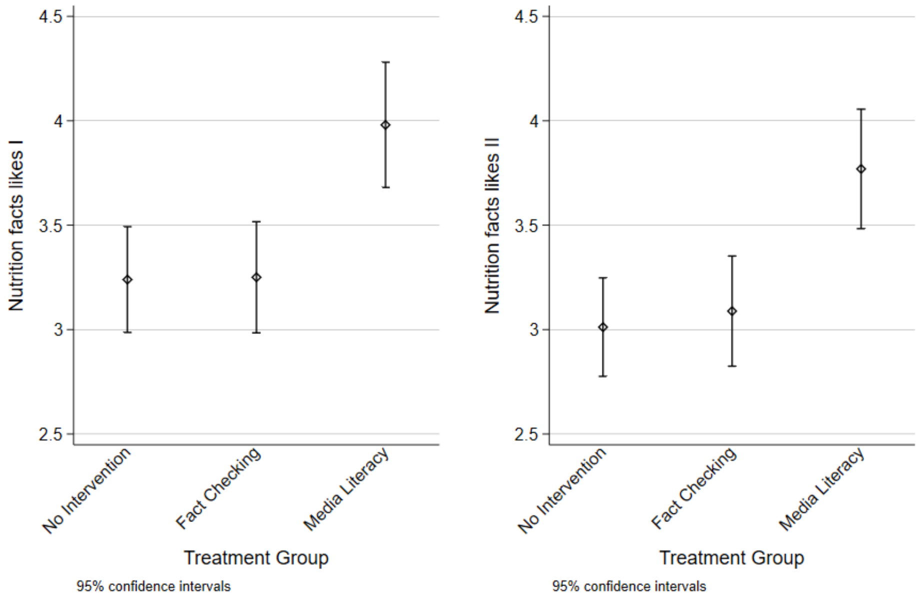


(b) Average number of untrustworthy elements for fakes on nutrition.

Figure B12: Average number of untrustworthy elements for fakes.

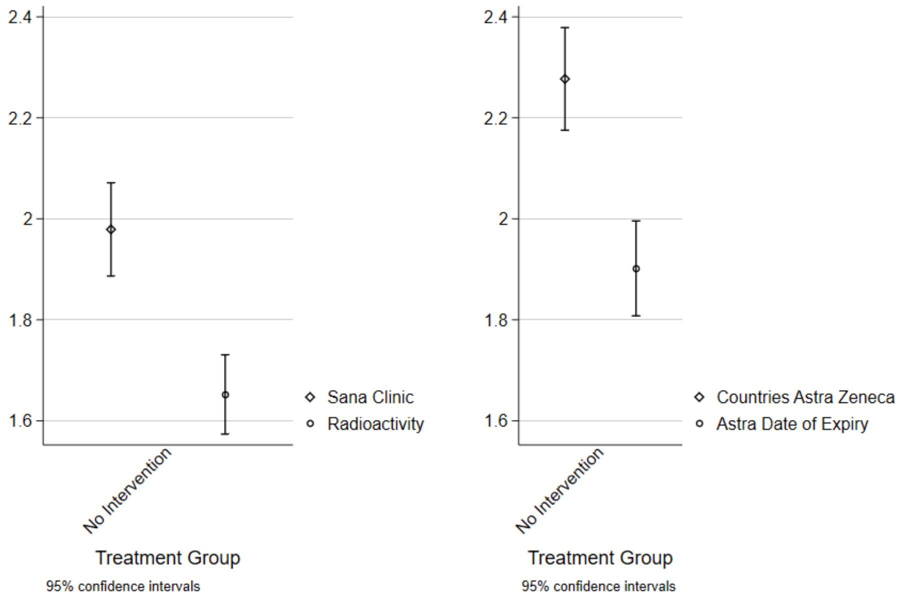


(a) Average number of trustworthy elements for facts on Corona vaccines.

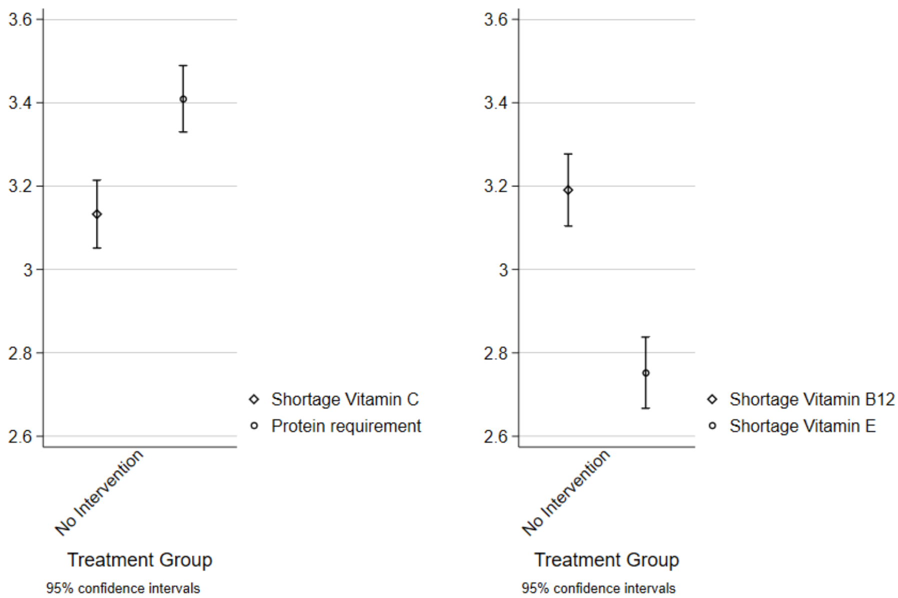


(b) Average number of trustworthy elements for facts on nutrition.

Figure B13: Average number of trustworthy elements for facts.

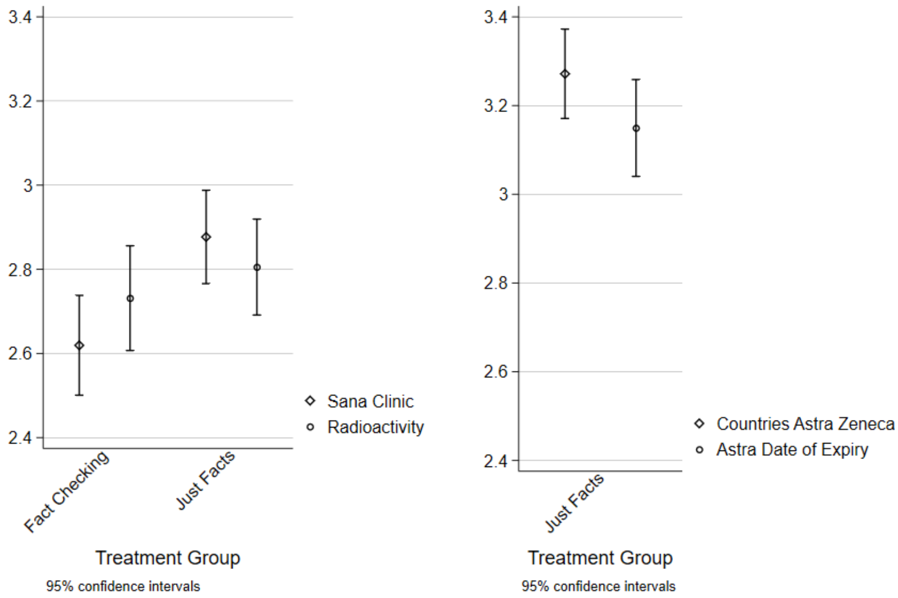


(a) Average credibility of fakes on Corona vaccines in NoIntervention group.

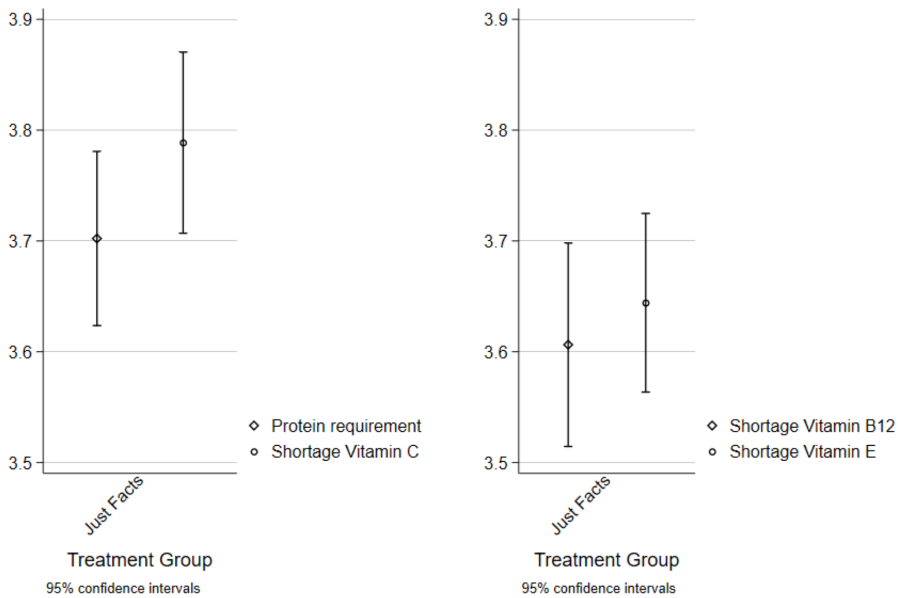


(b) Average credibility of fakes on nutrition in NoIntervention group.

Figure B14: Average credibility of fakes on a 5-point Likert Scale.



(a) Average credibility of fact-checks on Corona vaccines.



(b) Average credibility of fact-checks on nutrition.

Figure B15: Average credibility of fact-checks on a 5-point Likert Scale.

B.4 Omitted tables

Table B1: Balance table

	No Intervention		Fact-checking		Media Literacy		Passive Control		Just Facts	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
General:										
Age	40.741	[11.833]	40.542	[11.983]	40.165	[11.906]	40.241	[12.510]	41.075	[12.114]
Male	0.542	[0.499]	0.537	[0.499]	0.504	[0.500]	0.522	[0.500]	0.528	[0.500]
Vaccinated	0.788	[0.409]	0.835**	[0.371]	0.825*	[0.380]	0.830*	[0.376]	0.814	[0.389]
Supplement intake	0.450	[0.498]	0.445	[0.497]	0.434	[0.496]	0.493	[0.500]	0.432	[0.496]
Prior knowledge:										
log dist. current events	4.144	[1.122]	4.142	[1.148]	4.174	[1.168]	4.143	[1.149]	4.094	[1.146]
log dist. health	-0.021	[0.332]	-0.032	[0.342]	-0.015	[0.380]	-0.016	[0.349]	-0.037	[0.346]
log dist. nutrition	0.542	[0.537]	0.515	[0.535]	0.540	[0.536]	0.555	[0.544]	0.536	[0.531]
Family status:										
Fam: Married	0.411	[0.492]	0.376	[0.485]	0.383	[0.487]	0.403	[0.491]	0.405	[0.491]
Fam: Common law marriage	0.123	[0.329]	0.130	[0.337]	0.140	[0.348]	0.118	[0.323]	0.126	[0.332]
Fam: Unmarried	0.466	[0.499]	0.494	[0.500]	0.476	[0.500]	0.480	[0.500]	0.468	[0.499]
Household earnings:										
HH earnings < 1000	0.105	[0.307]	0.104	[0.305]	0.091	[0.288]	0.106	[0.309]	0.088	[0.284]
HH earnings [1000,1999]	0.222	[0.416]	0.189	[0.392]	0.207	[0.406]	0.224	[0.417]	0.261	[0.439]
HH earnings [2000,2999]	0.254	[0.436]	0.265	[0.442]	0.258	[0.438]	0.218	[0.413]	0.239	[0.427]
HH earnings [3000,3999]	0.162	[0.369]	0.170	[0.376]	0.170	[0.376]	0.167	[0.373]	0.201*	[0.401]
HH earnings > 4000	0.193	[0.395]	0.208	[0.406]	0.176	[0.381]	0.178	[0.383]	0.168	[0.374]
HH earnings n.s.	0.065	[0.246]	0.064	[0.245]	0.098**	[0.297]	0.106***	[0.309]	0.043*	[0.203]
Education:										
Education: no graduation	0.002	[0.040]	0.010*	[0.099]	0.007	[0.081]	0.005	[0.070]	0.002	[0.041]
Education: CSE (cat 1)	0.123	[0.329]	0.076***	[0.265]	0.088**	[0.284]	0.118	[0.323]	0.100	[0.300]
Education: CSE (cat2)	0.325	[0.469]	0.315	[0.465]	0.334	[0.472]	0.337	[0.473]	0.352	[0.478]
Education: high school	0.286	[0.452]	0.292	[0.455]	0.302	[0.459]	0.273	[0.446]	0.281	[0.450]
Education: college	0.264	[0.441]	0.308*	[0.462]	0.269	[0.444]	0.267	[0.443]	0.266	[0.442]
Personality traits:										
Big 5: conscientiousness	5.440	[1.086]	5.340	[1.118]	5.377	[1.111]	5.349	[1.118]	5.401	[1.114]
Big 5: extroversion	4.360	[1.331]	4.443	[1.345]	4.375	[1.344]	4.378	[1.306]	4.386	[1.294]
Big 5: tolerance	5.059	[1.073]	4.938**	[1.087]	5.097	[1.088]	5.050	[1.101]	5.035	[1.073]
Big 5: openness	4.563	[1.210]	4.462	[1.260]	4.528	[1.213]	4.556	[1.312]	4.539	[1.213]
Big 5: neuroticism	3.965	[1.350]	3.962	[1.322]	3.934	[1.257]	3.989	[1.340]	3.882	[1.309]
Party preferences:										
Vote: AfD	0.095	[0.294]	0.092	[0.290]	0.093	[0.291]	0.085	[0.279]	0.128*	[0.334]
Vote: CDU/CSU	0.173	[0.379]	0.157	[0.364]	0.153	[0.361]	0.164	[0.370]	0.128**	[0.334]
Vote: FDP	0.105	[0.307]	0.096	[0.294]	0.114	[0.318]	0.108	[0.311]	0.136*	[0.343]
Vote: Greens	0.176	[0.381]	0.199	[0.400]	0.176	[0.381]	0.195	[0.396]	0.173	[0.378]
Vote: Left	0.063	[0.243]	0.082	[0.275]	0.073	[0.261]	0.064	[0.245]	0.063	[0.243]
Vote: SPD	0.181	[0.386]	0.189	[0.392]	0.179	[0.384]	0.170	[0.376]	0.211	[0.408]
Vote: Other	0.206	[0.404]	0.185	[0.388]	0.210	[0.408]	0.214	[0.411]	0.161**	[0.368]
State of residence:										
State: Baden-Württ.	0.104	[0.305]	0.092	[0.290]	0.111	[0.314]	0.100	[0.300]	0.100	[0.300]
State: Bayern	0.146	[0.353]	0.145	[0.352]	0.148	[0.356]	0.162	[0.369]	0.143	[0.350]
State: Berlin	0.066	[0.249]	0.054	[0.227]	0.065	[0.247]	0.065	[0.248]	0.063	[0.243]
State: Brandenburg	0.026	[0.159]	0.028	[0.165]	0.036	[0.186]	0.029	[0.169]	0.033	[0.179]
State: Bremen	0.008	[0.090]	0.012	[0.107]	0.013	[0.114]	0.010	[0.099]	0.007	[0.081]
State: Hamburg	0.040	[0.197]	0.035	[0.183]	0.029	[0.169]	0.025	[0.155]	0.038	[0.192]
State: Hessen	0.084	[0.278]	0.091	[0.287]	0.067	[0.250]	0.088	[0.284]	0.078	[0.269]
State: Mecklenburg-Vorp.	0.023	[0.149]	0.016	[0.127]	0.013	[0.114]	0.011	[0.107]	0.022	[0.145]
State: Niedersachsen	0.087	[0.283]	0.112	[0.316]	0.095	[0.293]	0.077	[0.267]	0.081	[0.274]
State: Nordrhein-Westf.	0.204	[0.403]	0.224	[0.417]	0.192	[0.395]	0.221	[0.415]	0.211	[0.408]
State: Rheinland-Pfalz	0.052	[0.222]	0.058	[0.233]	0.060	[0.238]	0.043	[0.202]	0.033	[0.179]
State: Saarland	0.013	[0.113]	0.005	[0.070]	0.016	[0.127]	0.008	[0.090]	0.013	[0.115]
State: Sachsen	0.065	[0.246]	0.053	[0.224]	0.057	[0.232]	0.069	[0.253]	0.071	[0.258]
State: Sachsen-Anhalt	0.031	[0.173]	0.025	[0.155]	0.021	[0.144]	0.025	[0.155]	0.028	[0.166]
State: Schleswig-Holstein	0.034	[0.181]	0.025	[0.155]	0.044	[0.205]	0.049	[0.216]	0.050	[0.218]
State: Thüringen	0.018	[0.132]	0.026	[0.160]	0.031	[0.173]	0.018	[0.133]	0.028	[0.166]
N	618		607		613		611		602	

Notes: Table B1 displays the mean values and standard deviations of all our control variables for each treatment group. We also conducted t-tests on the difference in means between the NoIntervention and each of the other treatment groups respectively: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B2: Summary statistics of our dependent variables

Variable	Mean	Std. Dev.	Min.	Max.	N
Credibility:					
Prob. very credible/credible/indecisive – Corona vaccine fakes I	0.238	0.426	0	1	1834
Prob. very credible/credible/indecisive – Nutrition fakes I	0.876	0.329	0	1	1836
Prob. very credible/credible/indecisive – Corona vaccine fakes II	0.382	0.486	0	1	1533
Prob. very credible/credible/indecisive – Nutrition fakes II	0.798	0.401	0	1	1533
Prob. very credible/credible/indecisive – Corona vaccine facts I	0.861	0.346	0	1	2438
Prob. very credible/credible/indecisive – Nutrition facts I	0.946	0.226	0	1	2439
Prob. very credible/credible/indecisive – Corona vaccine facts II	0.867	0.339	0	1	2006
Prob. very credible/credible/indecisive – Nutrition facts II	0.956	0.206	0	1	2008
Factual knowledge:					
Distance to truth in SD – Corona vaccine fakes I	-0.002	0.802	-0.843	2.178	3051
Distance to truth in SD – Nutrition fakes I	0.001	0.727	-1.014	2.894	3051
Distance to truth in SD – Corona vaccine fakes II	-0.001	0.816	-0.907	1.876	2525
Distance to truth in SD – Nutrition fakes II	-0.004	0.74	-1.048	2.894	2525
Distance to truth in SD – Corona vaccine facts I	0.002	0.803	-0.676	2.965	3051
Distance to truth in SD – Nutrition facts I	0.002	0.776	-0.567	2.991	3051
Distance to truth in SD – Corona vaccine facts II	-0.001	0.734	-0.563	2.718	2525
Distance to truth in SD – Nutrition facts II	0.001	0.808	-0.793	1.789	2525
Attitudes:					
Prob. very likely/likely/indecisive – Covid vaccination I	0.813	0.39	0	1	3051
Prob. very likely/likely/indecisive – Dietary supplements I	0.548	0.498	0	1	3051
Prob. very likely/likely/indecisive – Covid vaccination II	0.781	0.413	0	1	2525
Prob. very likely/likely/indecisive – Dietary supplements II	0.553	0.497	0	1	2525
Trustworthy and untrustworthy elements:					
Corona vaccine fakes no. untrustworthy elements I	3.656	3.627	0	21	1838
Nutrition fakes no. untrustworthy elements I	1.466	1.888	0	13	1838
Corona vaccine fakes no. untrustworthy elements II	4.855	4.574	0	26	1546
Nutrition fakes no. untrustworthy elements II	1.898	2.467	0	16	1546
Corona vaccine facts no. trustworthy elements I	2.985	3.291	0	21	1838
Nutrition facts no. trustworthy elements I	3.386	3.377	0	19	2440
Corona vaccine facts no. trustworthy elements II	2.448	2.428	0	14	2028
Nutrition facts no. trustworthy elements II	3.214	3.076	0	20	2028
Corona vaccine fakes no. trustworthy elements I	0.671	1.419	0	17	1838
Nutrition fakes no. trustworthy elements I	1.711	1.958	0	12	1838
Corona vaccine fakes no. trustworthy elements I	1.42	1.998	0	15	1546
Nutrition fakes no. trustworthy elements II	1.25	1.87	0	15	1546
Corona vaccine facts no. untrustworthy elements I	1.039	1.376	0	11	1838
Corona vaccine facts no. untrustworthy elements II	0.707	0.965	0	7	2028
Nutrition facts no. untrustworthy elements I	0.74	1.356	0	10	2440
Nutrition facts no. untrustworthy elements II	0.716	1.424	0	11	2028

Table B3: Credibility of facts

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.012	-0.016	-0.003	-0.005	-0.009	-0.005	-0.021	-0.023
	[0.020]	[0.020]	[0.013]	[0.013]	[0.022]	[0.022]	[0.014]	[0.014]
p-value	(0.526)	(0.427)	(0.841)	(0.684)	(0.687)	(0.815)	(0.121)	(0.092)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,224	1,224	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	-0.006	-0.010	0.008	0.005	0.014	0.016	0.002	0.001
	[0.019]	[0.020]	[0.013]	[0.013]	[0.021]	[0.021]	[0.012]	[0.012]
p-value	(0.759)	(0.606)	(0.546)	(0.720)	(0.510)	(0.436)	(0.861)	(0.937)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,018	1,018	1,020	1,020
Baseline: No Intervention								
Mean DV	0.869	0.869	0.943	0.943	0.866	0.866	0.961	0.961
Std.Dev. DV	0.338	0.338	0.231	0.231	0.341	0.341	0.194	0.194

Notes: Table B3 presents the OLS coefficients of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The outcome is a dummy variable equal to one if participant i perceives the **facts** on Corona vaccines and nutrition in Wave I and Wave II of the survey as *Very credible*, *Credible* or *Indecisive* on average. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, income, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B4: Credibility – truth discernment

	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Fact-checking								
Fact-checking * fake	-0.068	-0.068	-0.013	-0.013	-0.000	-0.000	-0.005	-0.005
	[0.032]	[0.032]	[0.019]	[0.019]	[0.035]	[0.036]	[0.026]	[0.026]
p-value	(0.034)	(0.035)	(0.504)	(0.501)	(0.994)	(0.994)	(0.851)	(0.852)
Fact-checking	-0.012	-0.009	-0.003	-0.001	-0.009	0.006	-0.021	-0.018
	[0.020]	[0.020]	[0.013]	[0.013]	[0.022]	[0.023]	[0.014]	[0.014]
p-value	(0.526)	(0.639)	(0.842)	(0.910)	(0.687)	(0.801)	(0.121)	(0.190)
fake	-0.570	-0.570	-0.042	-0.042	-0.463	-0.464	-0.130	-0.130
	[0.023]	[0.023]	[0.013]	[0.013]	[0.024]	[0.025]	[0.017]	[0.017]
p-value	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,224	1,224	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy * fake	-0.098	-0.098	-0.067	-0.067	-0.066	-0.066	-0.074	-0.074
	[0.031]	[0.031]	[0.020]	[0.020]	[0.035]	[0.035]	[0.026]	[0.027]
p-value	(0.002)	(0.002)	(0.001)	(0.001)	(0.057)	(0.057)	(0.005)	(0.005)
Media literacy	-0.006	-0.009	0.008	0.006	0.014	0.020	0.002	0.003
	[0.019]	[0.020]	[0.013]	[0.013]	[0.021]	[0.022]	[0.012]	[0.012]
p-value	(0.759)	(0.663)	(0.546)	(0.660)	(0.510)	(0.347)	(0.861)	(0.790)
fake	-0.570	-0.570	-0.042	-0.042	-0.463	-0.464	-0.130	-0.130
	[0.023]	[0.023]	[0.013]	[0.013]	[0.025]	[0.025]	[0.017]	[0.017]
p-value	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,018	1,018	1,020	1,020
Baseline: No Intervention								
Mean DV	0.584	0.584	0.922	0.922	0.635	0.635	0.895	0.895
Std.Dev. DV	0.493	0.493	0.268	0.268	0.482	0.482	0.305	0.305

Notes: All estimates are OLS estimates. The dependent variable is a dummy equal to one if participant i considered the fakes or facts on Corona vaccines or nutrition in Wave I or Wave II of the survey on average as *Very credible*, *Credible* or *Indecisive*. Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively. Robust standard errors in squared parentheses, p-values in round parentheses. Standard errors are clustered on the participant level; N corresponds to the number of clusters. Control variables include age, gender, family status, income, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B5: Distance to truth on facts

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	0.059	0.059	-0.015	-0.005	0.056	0.062	-0.003	0.010
	[0.045]	[0.044]	[0.042]	[0.042]	[0.041]	[0.041]	[0.045]	[0.045]
p-value	(0.193)	(0.178)	(0.730)	(0.902)	(0.171)	(0.124)	(0.951)	(0.819)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	0.009	0.002	0.021	0.022	0.032	0.043	-0.005	0.013
	[0.044]	[0.042]	[0.042]	[0.040]	[0.043]	[0.042]	[0.045]	[0.044]
p-value	(0.831)	(0.971)	(0.613)	(0.581)	(0.447)	(0.305)	(0.908)	(0.761)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020
Baseline: No Intervention								
Mean DV	-0.094	-0.094	-0.100	-0.100	-0.180	-0.180	-0.206	-0.206
Std.Dev. DV	0.805	0.805	0.753	0.753	0.648	0.648	0.726	0.726

Notes: Table B5 presents OLS estimates for participants' distance to truth on topics that the **facts** on Corona vaccines and nutrition in Wave I and Wave II of the survey are dealing with. Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively. The dependent variable is equal to participant *i*'s average standardized distance to the correct answer. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B6: Distance to truth – truth discernment

	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Fact-checking								
Fact-checking * fake	-0.381	-0.381	-0.066	-0.066	-0.104	-0.104	0.023	0.023
	[0.071]	[0.071]	[0.049]	[0.049]	[0.075]	[0.075]	[0.059]	[0.060]
p-value	(0.000)	(0.000)	(0.181)	(0.181)	(0.164)	(0.167)	(0.698)	(0.700)
Fact-checking	0.059	0.060	-.0015	-0.010	0.056	0.058	-0.003	0.009
	[0.045]	[0.044]	[0.042]	[0.042]	[0.041]	[0.041]	[0.045]	[0.044]
p-value	(0.193)	(0.168)	(0.730)	(0.815)	(0.171)	(0.157)	(0.951)	(0.842)
fake	0.492	0.492	0.232	0.232	0.514	0.514	0.349	0.349
	[0.052]	[0.052]	[0.034]	[0.034]	[0.053]	[0.053]	[0.043]	[0.043]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,022	1,022	1,022	1,022
Panel B: Media literacy								
	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy * fake	-0.229	-0.229	-0.100	-0.100	-0.147	-0.147	-0.021	-0.021
	[0.071]	[0.071]	[0.050]	[0.050]	[0.075]	[0.076]	[0.059]	[0.060]
p-value	(0.001)	(0.001)	(0.045)	(0.046)	(0.052)	(0.053)	(0.721)	(0.723)
Media literacy	0.003	0.005	0.021	0.021	0.032	0.032	-0.005	-0.003
	[0.044]	[0.043]	[0.042]	[0.040]	[0.043]	[0.043]	[0.045]	[0.044]
p-value	(0.831)	(0.900)	(0.613)	(0.608)	(0.447)	(0.448)	(0.908)	(0.944)
fake	0.492	0.492	0.232	0.232	0.514	0.514	0.349	0.349
	[0.052]	[0.052]	[0.034]	[0.034]	[0.053]	[0.053]	[0.042]	[0.043]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020
Baseline: No Intervention								
Mean DV	0.152	0.152	0.016	0.016	0.077	0.077	-0.031	-0.031
Std.Dev. DV	0.859	0.859	0.705	0.705	0.795	0.795	0.678	0.678

Notes: All estimates are OLS estimates. The dependent variable is the average standardized distance to the true value that the fakes or facts on Corona vaccines or nutrition in Wave I or Wave II of the survey are covering. Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively. Robust standard errors in squared parentheses, p-values in round parentheses. Standard errors are clustered on the participant level; *N* corresponds to the number of clusters. Control variables include age, gender, family status, income, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B7: Credibility of fakes – Fact-checking vs. Media Literacy

	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media Literacy	-0.023	-0.031	-0.044	-0.048	-0.043	-0.052	-0.046	-0.056
	[0.023]	[0.023]	[0.020]	[0.019]	[0.030]	[0.029]	[0.026]	[0.026]
p-value	(0.320)	(0.024)	(0.024)	(0.013)	(0.150)	(0.077)	(0.076)	(0.030)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,216	1,216	1,218	1,218	1,024	1,024	1,024	1,024
Baseline: Fact-checking								
Mean DV	0.219	0.219	0.886	0.886	0.394	0.394	0.886	0.886
Std.Dev. DV	0.414	0.414	0.318	0.318	0.489	0.489	0.318	0.318

Notes: Table B7 shows the OLS estimates of comparing the FactChecking to the MediaLiteracy group. The dependent variable is a dummy equal to one if participant i perceives the **fakes** on Corona vaccines and nutrition in Wave I and in Wave II of the survey on average as *Very credible*, *Credible* or *Indecisive*. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B8: Distance to truth on topics covered by fakes – Fact-checking vs. Media Literacy

	Wave I				Wave II			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media Literacy	0.102	0.101	0.002	0.001	-0.066	-0.067	-0.046	-0.065
	[0.047]	[0.048]	[0.036]	[0.035]	[0.054]	[0.053]	[0.040]	[0.039]
p-value	(0.030)	(0.034)	(0.957)	(0.983)	(0.216)	(0.207)	(0.242)	(0.097)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,220	1,220	1,220	1,220	1,024	1,024	1,024	1,024
Baseline: Fact-checking								
Mean DV	0.076	0.076	0.051	0.051	0.286	0.286	0.164	0.164
Std.Dev. DV	0.818	0.818	0.604	0.604	0.870	0.870	0.638	0.638

Notes: Table B8 compares distance to truth on topics that the Corona vaccine and nutrition **fakes** are dealing with between participants from the FactChecking and the MediaLiteracy group. All estimates are OLS estimates. The dependent variable is equal to participant i's average standardized distance to the correct answer. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B9: Attitudes towards Corona vaccination and the intake of dietary supplements – Fact-checking vs. Media Literacy

	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Media Literacy	0.007	0.013	0.013	-0.040	-0.042	0.007	0.015	0.008	-0.033	-0.039
	[0.021]	[0.020]	[0.018]	[0.029]	[0.029]	[0.025]	[0.024]	[0.021]	[0.031]	[0.031]
p-value	(0.759)	(0.530)	(0.462)	(0.165)	(0.148)	(0.778)	(0.547)	(0.717)	(0.289)	(0.205)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,220	1,220	1,220	1,220	1,220	1,024	1,024	1,024	1,024	1,024
Baseline: Fact-checking										
Mean DV	0.830	0.830	0.830	0.555	0.555	0.797	0.797	0.797	0.565	0.565
Std.Dev. DV	0.376	0.376	0.376	0.497	0.497	0.402	0.402	0.402	0.496	0.496

Notes: Table B9 presents the OLS estimates of comparing the FactChecking to the MediaLiteracy group. The dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19, or *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the near future. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 8 ("yes +"), we also control for participants' Corona vaccination status.

Table B10: Heterogeneity in baseline beliefs on nutrition – Fact-checking

Panel A: Fact-checking – AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Suppl.	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.019	-0.054	0.091	-0.029	-0.107	0.037
	[0.067]	[0.156]	[0.099]	[0.075]	[0.108]	[0.117]
p-value	(0.783)	(0.728)	(0.361)	(0.704)	(0.329)	(0.755)
Controls	yes	yes	yes	yes	yes	yes
N	114	115	115	99	99	99
Baseline: No Intervention						
Mean DV	0.881	0.221	0.576	0.885	0.273	0.596
Std.Dev. DV	0.326	0.663	0.498	0.323	0.614	0.495
Panel B: Fact-checking – non-AfD supporters						
	Wave I			**Wave II**		
	Cred.	Dist.	Suppl.	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.009	-0.073	-0.003	-0.017	0.060	0.023
	[0.018]	[0.036]	[0.030]	[0.026]	[0.041]	[0.033]
p-value	(0.631)	(0.043)	(0.907)	(0.523)	(0.144)	(0.491)
Controls	yes	yes	yes	yes	yes	yes
N	1,109	1,110	1,110	923	923	923
Baseline: No Intervention						
Mean DV	0.903	0.122	0.555	0.825	0.129	0.549
Std.Dev. DV	0.296	0.629	0.497	0.380	0.572	0.498

Notes: Table B10 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) for our **Fact-checking** intervention. In columns 1 and 4, the dependent variable is a dummy equal to one if participant i perceives the fakes on nutrition as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant i's average standardized distance to the correct answer. In columns 3 and 6, the dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the near future. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B11: Heterogeneity in baseline beliefs on nutrition – Media literacy

Panel A: Media literacy – AfD supporters	**Wave I**			**Wave II**		
	Cred.	Dist.	Suppl.	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	0.043	0.044	0.006	-0.012	-0.156	0.001
	[0.063]	[0.152]	[0.101]	[0.068]	[0.153]	[0.100]
p-value	(0.495)	(0.774)	(0.952)	(0.866)	(0.309)	(0.990)
Controls	yes	yes	yes	yes	yes	yes
N	116	116	116	101	101	101
Baseline: No Intervention						
Mean DV	0.881	0.221	0.576	0.885	0.273	0.596
Std.Dev. DV	0.326	0.663	0.498	0.323	0.614	0.495

Panel B: Media literacy – non-AfD supporters	**Wave I**			**Wave II**		
	Cred.	Dist.	Suppl.	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	-0.067	-0.083	-0.033	-0.069	-0.027	-0.003
	[0.020]	[0.037]	[0.030]	[0.026]	[0.038]	[0.033]
p-value	(0.001)	(0.026)	(0.268)	(0.009)	(0.481)	(0.937)
Controls	yes	yes	yes	yes	yes	yes
Mean DV	0.869	0.079	0.537	0.788	0.122	0.542
Std.Dev. DV	0.337	0.632	0.499	0.409	0.605	0.499
N	1,115	1,115	1,115	919	919	919
Baseline: No Intervention						
Mean DV	0.903	0.122	0.555	0.825	0.129	0.549
Std.Dev. DV	0.296	0.629	0.497	0.380	0.572	0.498

Notes: Table B11 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) for our **Media literacy** intervention. In columns 1 and 4, the dependent variable is a dummy equal to one if participant i perceives the fakes on and nutrition as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant i's average standardized distance to the correct answer. In columns 3 and 6, the dependent variable is a dummy equal to one if participant i states to be *Very likely*, or *Likely* or *Indecisive* to consume dietary supplements in the near future. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B12: Persuasion rates credibility of fakes on Corona vaccines

Panel A: Fact-checking						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	-0.071	-0.166	-0.055	0.016	-0.221	0.046
Share to be persuaded	0.299	0.508	0.277	0.403	0.788	0.359
Persuasion rate	-0.237	-0.327	-0.199	0.040	-0.280	0.128
Panel B: Media literacy						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	-0.105	-0.134	-0.096	-0.042	-0.162	-0.027
Share to be persuaded	0.299	0.508	0.277	0.403	0.788	0.359
Persuasion rate	-0.351	-0.264	-0.347	-0.104	-0.206	-0.075

Notes: Table B12 displays the persuasion rates for our fact-checking (Panel A) and media literacy interventions (Panel B) for Wave I and Wave II of the survey, respectively. Columns 1 and 4 consider all participants in the NoIntervention, FactChecking, and MediaLiteracy groups. Columns 2 and 5 consider only AfD supporters, columns 3 and 6 only non-AfD supporters.

Table B13: Persuasion rates credibility of fakes on nutrition

Panel A: Fact-checking						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	-0.01	-0.019	-0.009	-0.018	-0.029	-0.017
Share to be persuaded	0.901	0.881	0.903	0.831	0.885	0.825
Persuasion rate	-0.011	-0.022	-0.010	-0.022	-0.033	-0.021
Panel B: Media literacy						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	-0.061	0.043	-0.067	-0.068	-0.012	-0.069
Share to be persuaded	0.901	0.881	0.903	0.831	0.885	0.825
Persuasion rate	-0.068	0.049	-0.074	-0.082	-0.014	-0.084

Notes: Table B13 displays the persuasion rates for our fact-checking (Panel A) and media literacy interventions (Panel B) for Wave I and Wave II of the survey, respectively. Columns 1 and 4 consider all participants in the NoIntervention, FactChecking, and MediaLiteracy groups. Columns 2 and 5 consider only AfD supporters, columns 3 and 6 only non-AfD supporters.

Table B14: Persuasion rates dietary supplements

Panel A: Fact-checking						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	0.003	0.091	-0.003	0.021	0.037	0.023
Share to be persuaded	0.557	0.576	0.555	0.554	0.596	0.549
Persuasion rate	0.005	0.158	-0.005	0.038	0.062	0.042
Panel B: Media literacy						
	Wave I			**Wave II**		
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate	-0.037	0.006	-0.033	-0.012	-0.039	-0.008
Share to be persuaded	0.557	0.576	0.555	0.554	0.596	0.549
Persuasion rate	-0.066	0.010	-0.059	-0.022	-0.065	-0.015

Notes: Table B14 displays the persuasion rates for our fact-checking (Panel A) and media literacy interventions (Panel B) for Wave I and Wave II of the survey, respectively. Columns 1 and 4 consider all participants in the NoIntervention, FactChecking, and MediaLiteracy groups. Columns 2 and 5 consider only AfD supporters, columns 3 and 6 only non-AfD supporters.

Table B15: Absolute number of trustworthy elements considered in fakes

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	0.035	0.041	-0.060	-0.046	-0.214	-0.186	0.001	0.034
	[0.082]	[0.083]	[0.108]	[0.109]	[0.120]	[0.123]	[0.117]	[0.119]
p-value	(0.668)	(0.617)	(0.578)	(0.671)	(0.076)	(0.131)	(0.992)	(0.773)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,030	1,030	1,030	1,030
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	-0.006	-0.031	0.128	0.098	-0.113	-0.128	0.032	0.018
	[0.082]	[0.082]	[0.114]	[0.113]	[0.131]	[0.134]	[0.113]	[0.111]
p-value	(0.941)	(0.706)	(0.264)	(0.389)	(0.388)	(0.339)	(0.776)	(0.873)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026
Baseline: No Intervention								
Mean DV	0.662	0.662	1.688	1.688	1.529	1.529	1.239	1.239
Std.Dev. DV	1.468	1.468	1.918	1.918	2.058	2.058	1.751	1.751

Notes: Table B15 compares the absolute number of trustworthy elements considered in fakes on Corona vaccines and dietary supplements for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B) in Wave I and Wave II of the survey, respectively. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B16: Absolute number of untrustworthy elements considered in facts

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	-0.156	-0.146	-0.126	-0.140	0.027	0.026	-0.065	-0.081
	[0.079]	[0.078]	[0.070]	[0.070]	[0.059]	[0.060]	[0.081]	[0.078]
p-value	(0.048)	(0.064)	(0.072)	(0.045)	(0.647)	(0.660)	(0.423)	(0.301)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,225	1,225	1,225	1,225	1,030	1,030	1,030	1,030
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	-0.025	-0.041	0.322	0.327	-0.043	-0.055	0.235	0.226
	[0.081]	[0.082]	[0.086]	[0.085]	[0.059]	[0.061]	[0.094]	[0.093]
p-value	(0.755)	(0.621)	(0.000)	(0.000)	(0.467)	(0.366)	(0.013)	(0.016)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026
Baseline: No Intervention								
Mean DV	1.099	1.099	0.696	0.696	0.704	0.704	0.665	0.665
Std.Dev. DV	1.491	1.491	1.315	1.315	0.915	0.915	1.366	1.366

Notes: Table B16 compares the absolute number of untrustworthy elements considered in facts on Corona vaccines and dietary supplements for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B) in Wave I and Wave II of the survey, respectively. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B17: Absolute number "verified account" labels considered as trustworthy

Panel A: Fact-checking				
	Corona		Nutrition	
	(1)	(2)	(3)	(4)
Fact-checking	-0.001	-0.003	0.002	0.001
	[0.017]	[0.018]	[0.011]	[0.012]
p-value	(0.973)	(0.874)	(0.835)	(0.926)
Controls	no	yes	no	yes
N	1,225	1,225	1,225	1,225
Panel B: Media literacy				
	Corona		Nutrition	
	(1)	(2)	(3)	(4)
Media literacy	0.097	0.089	0.059	0.054
	[0.024]	[0.023]	[0.014]	[0.014]
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes
N	1,231	1,231	1,231	1,231
Baseline: No Intervention				
Mean DV	0.058	0.058	0.040	0.040
Std.Dev. DV	0.306	0.306	0.197	0.197

Notes: Table B17 compares the absolute number "verified account labels" considered as trustworthy in facts on Corona vaccines and nutrition for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B). All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B18: Distance to truth on topics covered by fakes – Comparison to PassiveControl

Panel A: Fact-checking								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact-checking	0.512	0.533	0.022	0.053	0.757	0.764	0.240	0.259
	[0.041]	[0.040]	[0.043]	[0.042]	[0.043]	[0.043]	[0.046]	[0.047]
p-value	(0.000)	(0.000)	(0.615)	(0.210)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,218	1,218	1,218	1,218	1,030	1,030	1,030	1,030
Panel B: Media literacy								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Media literacy	0.614	0.604	0.024	0.039	0.691	0.676	0.194	0.198
	[0.041]	[0.040]	[0.044]	[0.044]	[0.042]	[0.041]	[0.046]	[0.046]
p-value	(0.000)	(0.000)	(0.593)	(0.373)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,224	1,224	1,224	1,224	1,028	1,028	1,028	1,028
Panel C: No Intervention								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No Intervention	0.834	0.840	0.102	0.109	0.805	0.808	0.220	0.222
	[0.041]	[0.041]	[0.043]	[0.043]	[0.041]	[0.042]	[0.045]	[0.045]
p-value	(0.000)	(0.000)	(0.020)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,229	1,229	1,229	1,229	1,026	1,026	1,026	1,026
Baseline: Passive Control								
Mean DV	-0.233	-0.233	-0.264	-0.264	-0.397	-0.397	-0.396	-0.396
Std.Dev. DV	0.600	0.600	0.772	0.772	0.593	0.593	0.840	0.840

Notes: Table B18 presents OLS estimates for participants' distance to truth on topics that the **fakes** on Corona vaccines and nutrition in Wave I and Wave II of the survey are dealing with. Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the PassiveControl group, respectively. The dependent variable is equal to participant *i*'s average standardized distance to the correct answer. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B19: Attitudes towards Corona vaccination and the intake of dietary supplements – Comparison to PassiveControl

Panel A: Fact-checking										
	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Fact-checking	-0.006	-0.014	-0.011	-0.034	-0.038	0.012	0.007	0.007	-0.003	0.003
	[0.021]	[0.020]	[0.018]	[0.028]	[0.028]	[0.025]	[0.024]	[0.021]	[0.031]	[0.031]
p-value	(0.778)	(0.464)	(0.541)	(0.231)	(0.173)	(0.637)	(0.784)	(0.747)	(0.913)	(0.929)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,218	1,218	1,218	1,218	1,218	1,030	1,030	1,030	1,030	1,030
Panel B: Media literacy										
	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Media literacy	0.001	-0.000	0.001	-0.074	-0.075	0.019	0.023	0.015	-0.036	-0.033
	[0.021]	[0.021]	[0.019]	[0.028]	[0.028]	[0.025]	[0.024]	[0.022]	[0.031]	[0.031]
p-value	(0.980)	(0.981)	(0.972)	(0.009)	(0.007)	(0.451)	(0.338)	(0.493)	(0.242)	(0.282)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,224	1,224	1,224	1,224	1,224	1,028	1,028	1,028	1,028	1,028
Panel C: No Intervention										
	Wave I					**Wave II**				
	Corona vaccination			Supplements		Corona vaccination			Supplements	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
No Intervention	-0.053	-0.051	-0.032	-0.033	-0.035	-0.045	-0.038	-0.028	-0.015	-0.016
	[0.022]	[0.021]	[0.019]	[0.028]	[0.028]	[0.027]	[0.025]	[0.022]	[0.031]	[0.031]
p-value	(0.018)	(0.015)	(0.092)	(0.249)	(0.210)	(0.093)	(0.122)	(0.219)	(0.637)	(0.614)
Controls	no	yes	yes +	no	yes	no	yes	yes +	no	yes
N	1,229	1,229	1,229	1,229	1,229	1,026	1,026	1,026	1,026	1,026
Baseline: No Intervention										
Mean DV	0.836	0.836	0.836	0.589	0.589	0.785	0.785	0.785	0.569	0.569
Std.Dev. DV	0.370	0.370	0.370	0.492	0.492	0.411	0.411	0.411	0.496	0.496

Notes: Table B19 presents the OLS estimates of comparing the PassiveControl to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant *i* states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19, or *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the near future. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 8 ("yes +"), we also control for participants' Corona vaccination status.

Table B20: Distance to truth on topics covered by fakes – JustFacts group

Panel A: Comparison to NoIntervention group								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Just facts	-0.631	-0.641	-0.596	-0.588	-0.731	-0.729	-0.540	-0.535
	[0.042]	[0.042]	[0.061]	[0.061]	[0.046]	[0.047]	[0.046]	[0.047]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
Mean DV	0.398	0.398	0.212	0.212	0.333	0.333	0.143	0.143
Std.Dev. DV	0.841	0.841	0.769	0.769	0.844	0.844	0.577	0.577
N	1,220	1,220	1,220	1,220	984	984	984	984
Baseline: No Intervention								
Mean DV	0.398	0.398	0.212	0.212	0.333	0.333	0.143	0.143
Std.Dev. DV	0.841	0.841	0.769	0.769	0.844	0.844	0.577	0.577
Panel B: Comparison to PassiveControl group								
	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Just facts	0.202	0.222	-0.277	-0.239	0.074	0.080	-0.319	-0.302
	[0.034]	[0.033]	[0.069]	[0.067]	[0.033]	[0.034]	[0.053]	[0.055]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.024)	(0.019)	(0.000)	(0.000)
Controls	no	yes	no	yes	no	yes	no	yes
N	1,213	1,213	1,213	1,213	992	992	992	992
Baseline: No Intervention								
Mean DV	-0.436	-0.436	0.030	0.030	-0.471	-0.471	-0.077	-0.077
Std.Dev. DV	0.589	0.589	0.879	0.879	0.419	0.419	0.833	0.833

Notes: Table B20 compares distance to truth on topics that the Corona vaccine and nutrition **fakes** are dealing with between participants from the NoIntervention (Panel A) and the PassiveControl (Panel B) and the JustFacts group, respectively. All estimates are OLS estimates. The dependent variable is equal to participant i's average standardized distance to the correct answer. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B21: List experiments – Sample split

Panel A: Vaccination				
	Likely to get vaccinated		**Unlikely to get vaccinated**	
	Wave I	Wave II	Wave I	Wave II
	(1)	(2)	(3)	(4)
List experiment	7.06	0.00	52.71	35.69
N	2,480	1,973	571	552
Panel B: Supplements				
	Likely to take supplements		**Unikely to take supplements**	
	Wave I	Wave II	Wave I	Wave II
	(1)	(2)	(3)	(4)
List experiment	79.23	79.01	19.23	12.99
N	918	780	2,133	1,745

Notes: Panel A splits participants who directly report to be *Very likely*, *Likely* or *Indecisive* to get vaccinated against Covid-19 in the main experiment from those who did not and displays the respective indirectly elicited proportions from the list experiments for each of those subsamples. Analogously, Panel B splits participants who directly report to be *Very likely*, *Likely* or *Indecisive* to consume dietary supplements in the main experiment from those who did not and displays the respective proportions from the list experiments.

Table B22: 2SLS estimates for our main specifications

Panel A: Fact-checking												
	Wave I						**Wave II**					
	Corona			Nutrition			Corona			Nutrition		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Second stage												
$\widehat{D_i}$	-0.101	-0.458	0.025	-0.015	-0.115	0.005	0.023	-0.072	0.047	-0.026	0.043	0.030
	[0.034]	[0.070]	[0.026]	[0.025]	[0.049]	[0.040]	[0.042]	[0.077]	[0.031]	[0.034]	[0.054]	[0.044]
p-value	(0.003)	(0.000)	(0.332)	(0.558)	(0.020)	(0.907)	(0.578)	(0.344)	(0.122)	(0.453)	(0.420)	(0.497)
First stage												
Fact-checking	0.699	0.696	0.698	0.699	0.697	0.697	0.704	0.704	0.705	0.704	0.704	0.704
	[0.018]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]	[0.020]	[0.020]	[0.020]	[0.020]	[0.020]	[0.020]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F-statistic	1429.95	1409.5	1408.42	1421.1	1409.5	1409.5	1186.3	1186.3	1186.61	1186.3	1186.3	1186.3
Controls	yes	yes	yes +	yes	yes	yes	yes	yes	yes +	yes	yes	yes
N	1,221	1,225	1,225	1,223	1,225	1,225	1,022	1,022	1,022	1,022	1,022	1,022
Panel B: Media literacy												
	Wave I						**Wave II**					
	Corona			Nutrition			Corona			Nutrition		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.	Cred.	Dist.	Vaccine	Cred.	Dist.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Second stage												
$\widehat{D_i}$	-0.142	-0.296	0.046	-0.082	-0.109	-0.051	-0.058	-0.189	0.065	-0.093	-0.056	-0.017
	[0.031]	[0.065]	[0.025]	[0.025]	[0.048]	[0.038]	[0.039]	[0.071]	[0.030]	[0.033]	[0.050]	[0.042]
p-value	(0.000)	(0.000)	(0.071)	(0.001)	(0.024)	(0.181)	(0.135)	(0.008)	(0.030)	(0.005)	(0.262)	(0.690)
First stage												
Media Literacy	0.740	0.740	0.741	0.740	0.740	0.740	0.731	0.731	0.733	0.731	0.731	0.731
	[0.017]	[0.017]	[0.017]	[0.017]	[0.017]	[0.017]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]	[0.019]
p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F-statistic	1799.59	1799.59	1797.85	1799.59	1799.59	1799.59	1411.51	1411.51	1412.8	1411.51	1411.51	1411.51
Controls	yes	yes	yes +	yes	yes	yes	yes	yes	yes +	yes	yes	yes
N	1,231	1,231	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020	1,020	1,020
Baseline: No Intervention												
Mean DV	0.299	0.398	0.783	0.901	0.132	0.557	0.403	0.334	0.741	0.831	0.143	0.554
Std.Dev. DV	0.458	0.841	0.412	0.299	0.632	0.497	0.491	0.844	0.439	0.375	0.577	0.498

Notes: Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition. In columns 3 and 9 ("yes +"), we also control for participants' Corona vaccination status.

Table B23: Heterogeneity in vaccination status – FactChecking

Panel A: Fact-checking – Fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.059	-0.323	0.018	0.037	-0.049	0.036
	[0.026]	[0.054]	[0.018]	[0.032]	[0.060]	[0.023]
p-value	(0.023)	(0.000)	(0.310)	(0.243)	(0.414)	(0.119)
Controls	yes	yes	yes	yes	yes	yes
N	991	994	994	835	835	835
Panel B: Fact-checking – Not fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Fact-checking	-0.056	-0.299	0.034	-0.051	-0.043	0.049
	[0.069]	[0.115]	[0.065]	[0.075]	[0.142]	[0.071]
p-value	(0.417)	(0.010)	(0.596)	(0.496)	(0.765)	(0.488)
Controls	yes	yes	yes	yes	yes	yes
N	230	231	231	187	187	187
Baseline: No Intervention						
Mean DV	0.299	0.398	0.783	0.403	0.334	0.741
Std.Dev. DV	0.458	0.841	0.412	0.491	0.844	0.439

Notes: Table B23 displays the effect heterogeneity between fully vaccinated (Panel A) and not fully vaccinated (Panel B) participants for our **Fact-checking** intervention. The NoIntervention group is the omitted category in all specifications. In columns 1 and 4, the dependent variable is a dummy equal to one if participant *i* perceives the fakes on Corona vaccines as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant *i*'s average standardized distance to the correct answer. In columns 3 and 9, the dependent variable is a dummy equal to one if participant *i* states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B24: Heterogeneity in vaccination status – MediaLiteracy

Panel A: Media literacy – Fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	-0.093	-0.211	0.011	-0.020	-0.187	0.041
	[0.025]	[0.054]	[0.018]	[0.031]	[0.058]	[0.023]
p-value	(0.000)	(0.000)	(0.558)	(0.530)	(0.001)	(0.077)
Controls	yes	yes	yes	yes	yes	yes
N	993	993	993	835	835	835
Panel B: Media literacy – Not fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Dist.	Vaccine	Cred.	Dist.	Vaccine
	(1)	(2)	(3)	(4)	(5)	(6)
Media literacy	-0.081	-0.292	0.119	-0.047	0.058	0.056
	[0.066]	[0.106]	[0.061]	[0.073]	[0.125]	[0.069]
p-value	(0.216)	(0.006)	(0.054)	(0.526)	(0.642)	(0.423)
Controls	yes	yes	yes	yes	yes	yes
N	238	238	238	185	185	185
Baseline: No Intervention						
Mean DV	0.299	0.398	0.783	0.403	0.334	0.741
Std.Dev. DV	0.458	0.841	0.412	0.491	0.844	0.439

Notes: Table B24 displays the effect heterogeneity between fully vaccinated (Panel A) and not fully vaccinated (Panel B) participants for our **media literacy** intervention. The NoIntervention group is the omitted category in all specifications. In columns 1 and 4, the dependent variable is a dummy equal to one if participant i perceives the fakes on Corona vaccines as *Very credible*, *Credible* or *Indecisive* on average. In columns 2 and 5, the dependent variable is equal to participant i's average standardized distance to the correct answer. In columns 3 and 9, the dependent variable is a dummy equal to one if participant i states to be *Very likely*, *Likely* or *Indecisive* to get vaccinated or boostered against Covid-19. All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

B.5 Follow-Up Experiment

B.5.1 Introduction

We run a follow-up experiment to provide additional evidence about the underlying mechanisms and examine the generalizability of our results to a different context. We aim to closely match the original study's experimental design, implementing changes only where it is needed to gain the desired insights. In particular, we change three main aspects compared to the original study:

1. We test the effectiveness of fact-checking and media literacy interventions in a different context. Instead of exposing participants to Covid-19-related fakes, we expose them to posts containing false information about environmental topics.[1] This new context allows us to better understand the external validity of our findings, particularly in a setting which is less linked to current events (such as the Covid-19 crises in 2020/2021) and likely to be of long-term relevance.

2. We randomize the order in which fact-checks and fakes are presented. This is done to assess whether our results change when fact-checks are shown before or after the corresponding fakes.

3. To analyse whether fact-checking more credible fakes enhances participants' critical evaluation of subsequent posts, we varied the exposure by presenting fakes with a higher mean baseline credibility, which were fact-checked before participants proceeded to a new topic without further intervention.

The following paragraphs describe the design of the follow-up experiment and its results.

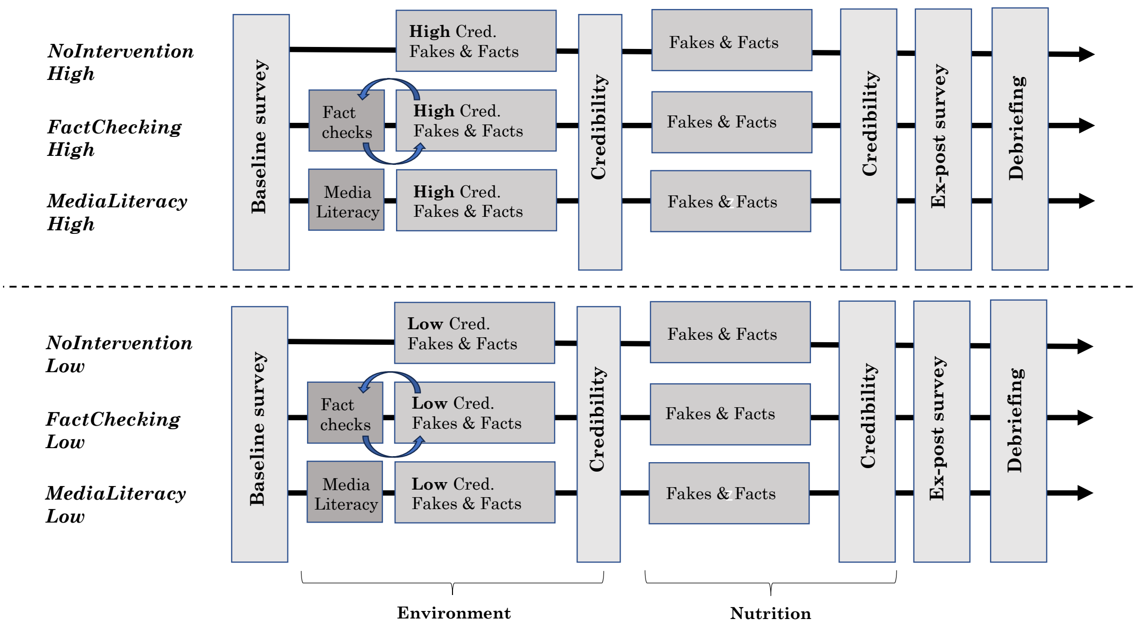
B.5.2 Experimental Design

B.5.2.1 Survey flow

All of the questions we aim to answer with the follow-up can be assessed with either a comparison of the NoIntervention group (control group) to the intervention groups (MediaLiteracy and FactChecking), or with a direct comparison of both intervention groups or within intervention groups. We therefore do not collect any new data on a PassiveControl or JustFacts group.

[1] For literature reviews on online misinformation about environmental topics see Lewandowsky (2021), Gundersen et al. (2022), and Erbaugh et al. (2024).

Figure B16: Experimental Procedures Follow-Up – Overview



Notes: Figure B16 gives an overview of the experimental procedures. The groups that were exposed to the high-credibility fakes are displayed above the dashed line, the others below. The blue arrows indicate the randomization of the order of fact-checks and fakes in the fact-checking group.

Because we also aim to cleanly identify potential differences between exposure to high and low credibility fakes a 2x3 design is necessary. The "high" and "low" NoIntervention groups are therefore identical except for the specific posts used during the first part of the survey, and similar for the "high" and "low" MediaLiteracy groups and the "high" and "low" FactChecking groups. In contrast to the original experiment, this follow-up includes posts that deal with environmental topics instead of Covid-19 vaccinations. To select those, we compiled a database of Facebook posts and selected the posts that were rated as most credible and least credible in a pre-study (see Appendix B.5.2.4 for details).

The second part of the survey, which aims to explore the effectiveness of the intervention in a different context, remains very similar to Wave I of the original experiment, utilizing the same posts about the topic *Nutrition*. Figure B16 provides a graphical overview of the experimental procedures of the follow-up. Participants receive a full debriefing at the end of their participation, including all fact-checks for the false posts they were exposed to.

B.5.2.2 Implementation

Again, the follow-up experiment was implemented with the survey software *Qualtrics*. As in the Wave I of our original experiment we recruited a minimum of 600 participants per group from the subject pool of the market research company bilendi (formerly respondi), which should enable us to identify treatment effects even if they are relatively small. The follow-up experiment was conducted online between July 9th and July 18th 2024 with N=4,003 subjects. The participants are representative of the German population in terms of age, gender and state of residency. Participants of our first study were excluded from participation. The median time for completing the survey was 19.95 minutes. The follow-up experiment and our hypothesis regarding the results have been pre-registered on the AEA RCT Registry with the ID 13629.[2]

B.5.2.3 Balance check

Our sample is representative for the German population in terms of age, gender and state. The six treatment groups are mostly balanced with respect to covariates, which indicates that the randomization has worked well. We test the pairwise balance of covariates using either t-tests or Chi squared tests. Out of the resulting 945 tests, 44 indicate statistically significant differences between two groups. Note that these slight imbalances can only be due to chance as we randomized by computer and there was close to no attrition. We address this issue by presenting results where we control for our set of covariates.

B.5.2.4 Selection of new fakes and facts

In contrast to the original experiment, the follow-up includes posts that deal with environmental topics instead of Covid-19 vaccinations. To select those, we compiled a database of Facebook posts that fulfilled the following criteria: The posts should be related to environmental topics and fakes should contain a false claim that has been flagged as false by an independent fact-checking organization.

We focus on these somewhat looser criteria compared to our original study, because we wanted to be able to collect a larger number of posts. In our first study, posts also had to contain a statistic and had to potentially provoke behavioral implications of participants (like getting vaccinated or taking dietary supplements). In the interest of learning more about the external validity of our findings we settled on the above described criteria and focus on the measurement of credibility as a main outcome. Accordingly, our measures for knowledge and behavioral attitudes cannot be computed in this follow-up.

In a pre-study, we evaluate the credibility of posts in the resulting database. To this end, 30 German-speaking participants, recruited via Prolific, rated the credibility of the nine fakes and ten facts on a Likert scale from 1 to 5. To obtain posts with

[2]The pre-registration can be accessed here: `https://www.socialscienceregistry.org/trials/13629`.

different credibility levels for this follow-up study we selected the two most credible and the two least credible posts for fakes and facts. The fake posts have been flagged as fake by the independent fact-checking organization *Correctiv.org*.

B.5.2.5 Variables

Credibility Our measure for credibility is exactly as in the original study. Participants are asked to rate the credibility of every posting they see on a 5-point-Likert scale. The responses are averaged across environment fakes, environment facts, nutrition fakes and nutrition facts (i.e., we compute four mean responses per participant). We then generate a dummy equal to 1 if participants rated the fakes or facts on average as *Very credible*, *Credible* or *Indecisive*.

Time spent with posts We measure how many seconds each participant spends on the screen viewing a post before being asked to evaluate its credibility. This figure is summed across all fact and fake posts to calculate the total viewing time per participant.

Control variables We elicit our set of control variables[3] in the same way as in our original study. We only implement a slight change in the question on prior knowledge of current events. In particular, we ask how many months remain until the next U.S. presidential election instead of asking how many days president Biden has been in office.

B.5.2.6 Analysis

For our baseline analysis we adhere as close as possible to our original empirical strategy. Hence, we test the effectiveness of the fact-checking and media literacy interventions using the regression specifications as described equation (3.1) in section 3.3.5. However, as there is only one outcome (*credibility*) and one survey wave, the equation reduces to:

$$cred_i = \beta_0 + \beta_1 TG_i + \beta_2 X_i + \varepsilon_i, \tag{B.1}$$

where $cred_i$ corresponds to the credibility outcome of participant i, TG_i denotes participant i's treatment group, and X_i is a vector of pre-registered control variables including age, gender, party preferences, religion, education, family status, household income, personality traits, state of residence, and prior knowledge on current events, health, and nutrition. The baseline category in TG_i is the NoIntervention group,

[3]Our standard control variables include age, gender, family status, income, education, "big 5", political preferences and prior knowledge on current events, health, and nutrition.

i.e., we compare participants who receive fakes and facts without further intervention to participants in each of the other treatment groups. To this end, we pool the full sample of participants of the follow-up experiment, not making a difference between participants who receive low credibility and those who receive high credibility posts.[4]

B.5.2.7 Hypotheses

We pre-registered our hypotheses for the experiment outcomes on the AEA RCT Registry under the ID 13629 (`https://www.socialscienceregistry.org/trials/13629`). Based on existing literature showing that fact-checking and media literacy interventions are effective beyond the context of the Covid-19-pandemic, we were confident that our interventions would be effective for fake posts about the environment. Furthermore, we hypothesized that both fact-checking and media literacy interventions would be more effective when participants were exposed to fakes that are easier to debunk (the "low-credibility fakes"). We, therefore, test the following pre-registered hypotheses:

H1 The fact-checking intervention reduces the credibility of the "fake news" as compared to participants without intervention.

H2 The media literacy intervention reduces the credibility of the "fake news" as compared to participants without intervention.

H3 The effects of both interventions are higher for fakes which are ex-ante rated as less credible.

We furthermore examine potential differences in the time participants spend on the posts to better understand the mechanisms through which the interventions work. This measure was not collected in the original study.

B.5.3 Results

We present our results in the following order: First, we conduct a manipulation check of the degree of credibility. Second, we proceed to demonstrate how our media literacy and fact-checking interventions influence the credibility of fakes in the context of posts related to the environment and nutrition. Third, we analyse how the order of fact-check and fake affects the effectiveness of fact-checking. Forth,

[4]We preregistered our baseline regression estimation as described here. See Appendix B.5.2.7 for details.

Table B25: Mean values of all outcomes per treatment group

Panel A: Pooled Sample						
	NoIntervention		**FactChecking**		**MediaLiteracy**	
	Environment	Nutrition	Environment	Nutrition	Environment	Nutrition
	(1)	(2)	(3)	(4)	(5)	(6)
Credibility	0.613	0.830	0.582	0.784	0.438	0.677
	(0.013)	(0.010)	(0.013)	(0.011)	(0.014)	(0.013)
Panel B: Credibility split						
	NoIntervention		**FactChecking**		**MediaLiteracy**	
	Environment	Nutrition	Environment	Nutrition	Environment	Nutrition
	(1)	(2)	(3)	(4)	(5)	(6)
High credibility	0.598	0.821	0.555	0.786	0.453	0.669
	(0.019)	(0.015)	(0.019)	(0.016)	(0.019)	(0.018)
Low credibility	0.629	0.840	0.608	0.783	0.433	0.684
	(0.018)	(0.014)	(0.019)	(0.016)	(0.019)	(0.018)

Notes: Table B25 shows the mean credibility values for each of our treatment groups. Panel A displays the full sample and Panel B splits the sample by the credibility of posts, as assessed in the pre-study, during the first part of the survey. Standard errors in parentheses.

we provide insights on the role of the level of credibility of fakes for the effectiveness of the different interventions. Fifth, we show additional evidence for the proposed mechanism by showing how the time spent with posts varies by group. Finally, we analyse the probability to fail an attention check for the different groups.

B.5.3.1 Manipulation Check regarding "High" and "Low" Credibility

A comparison between the "high" and "low" credibility groups can only be interpreted meaningfully, if the pre-study assessment of the fakes' credibility at least qualitatively matches the findings for the NoIntervention groups in the follow-up. However, this is not the case. The means of the perceived credibility of the posts across the "high" and the "low"NoIntervention groups are very similar. If anything, the mean in the "high" credibility group is slightly lower than the one in the "low"credibility group, which stems from one post being evaluated substantially differently by the participants in this follow-up compared to the pre-study. Table B25 presents mean outcomes per treatment group.

As the manipulation in terms of heterogeneity of credibility of the stimuli did not work as intended, we do not follow our pre-registered analysis plan and do not interpret differences between the groups who saw different posts as being caused by heterogeneity in the baseline credibility of these posts. All other planned analyses are possible in this setting. As the credibility of the fakes is similar for the "high" and the "low" groups, we pool these two groups for all other analyses. We provide a detailed description of the reasons for the failed manipulation check and an alternative analysis of the influence of the degree of credibility on the effectiveness of our interventions in Appendix B.5.3.5.

Table B26: Credibility of fakes

Panel A: Fact-checking				
	Environment		Nutrition	
	(1)	(2)	(3)	(4)
Fact-checking	-0.031	-0.033	-0.046	-0.046
	[0.018]	[0.018]	[0.015]	[0.015]
p-value	(0.101)	(0.074)	(0.002)	(0.002)
Controls	no	yes	no	yes
N	2,679	2,679	2,679	2,679
Panel B: Media literacy				
	Environment		Nutrition	
	(1)	(2)	(3)	(4)
Media literacy	-0.175	-0.171	-0.153	-0.148
	[0.019]	[0.018]	[0.016]	[0.016]
p-value	(0.000)	(0.000)	(0.000)	(0.000)
Controls	no	yes	no	yes
N	2,663	2,663	2,663	2,663
Baseline: No Intervention				
Mean DV	0.613	0.613	0.830	0.830
Std.Dev. DV	0.487	0.487	0.375	0.375

Notes: Table B26 shows the OLS estimates of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant i did perceive the **fakes** on environment and nutrition on average as *Very credible*, *Credible* or *Indecisive*. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

B.5.3.2 Effectiveness of Interventions with Posts about the Environment

Our data suggests that the intervention for both the FactChecking and the MediaLiteracy group are effective in a different context than the one from our original experiment: They help to reduce the credibility of fakes on environmental topics. Participants in the FactChecking group are 3.3 percentage points less likely to rate a post as *Very credible, Credible* or *Indecisive* ($p = 0.074$) than those in the NoIntervention group. This corresponds to a 6.7% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 5.4% of its mean value. Those in the MediaLiteracy group are 17.1 percentage points less likely to do so ($p < 0.001$) which corresponds to a 35.1% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 27.9% of its mean value.[5]

Relative to the original experiment, the coefficients for the FactChecking group are smaller and less precisely estimated, while for the MediaLiteracy group the absolute effect size appears substantially larger; however, as a percentage of the standard deviation and baseline mean, both remain in a similar range. These findings align with our hypotheses H1 and H2, indicating that both interventions lower the perceived credibility of fake posts compared to participants who did not receive an intervention. Detailed regression results are available in columns 1 and 2 in table B26.

B.5.3.3 Effectiveness of Interventions with Posts about Nutrition

Because we use the same fakes on nutrition as in the first wave of our original experiment, this follow-up allows us to compare these findings to the results on nutrition posts in our first experiment. Interestingly, the new data suggests that both the fact-checking and the media literacy intervention are effective in reducing the credibility of the nutrition fakes, however, the effect is substantially larger for the MediaLiteracy group. Participants in the FactChecking group are 4.6 percentage points less likely to rate the fakes as *Very credible, Credible* or *Indecisive* than participants in the NoIntervention group. The estimate is statistically significant at the 1%-level; the effect size corresponds to about 12.2% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 5.5% of its mean value.[6] Participants in the MediaLiteracy group

[5] A direct comparison between the two treatments indicates that participants of the MediaLiteracy group are 14 percentage points less likely to rate a post as *Very credible, Credible* or *Indecisive* ($p < 0.001$) than those of the FactChecking group.

[6] We elaborate further on this finding in Appendix B.5.3.5.

in contrast are about 15 percentage points less likely to do so. The estimate is statistically significant at the 1%-level; the effect size corresponds to about 40% a standard deviation in the dependent variable in the baseline NoIntervention group and to about 18% its mean value.[7]. Compared to the effects documented in our first study, these coefficients are larger and more precise, which increases our confidence in our main finding that the media literacy intervention likely raises users' skills and awareness to spot misinformation. Columns 3 and 4 of table B26 provide detailed regression results.

B.5.3.4 Position of the fact-check

In the previous experiment fact-checks were always displayed to the participants before seeing the actual fake post. In reality, platforms implement fact-checks both before and after posts with fake content. We randomized this order in the follow-up experiment to test whether the order of fact-check and fake influences the initial experiment's findings. Therefore, we use OLS to estimate the following regression equation within the FactChecking group

$$cred_i = \beta_0 + \beta_1 post_i + \beta_2 mix_i + \beta_3 X_i + \varepsilon_i, \tag{B.2}$$

where $cred_i$ corresponds to the credibility outcome of participant i, $post_i$ equals one if participant i saw the fact-check always after the fake, mix_i equals one if participant i saw it once before and once after, the baseline represents participants who always saw it before, and X_i is a vector of pre-registered control variables as in equation 5. Using the default cutoff for the dependent variable where it equals one when participant i perceives the fakes on average as *Very credible*, *Credible* or *Indecisive* the fact-checking intervention appears to be most effective when fact-checks are presented after the fake. Including controls the coefficient turns statistically significant on the 5%-level (see table B27). These results are not robust to a slightly stricter cutoff in the dependent variable. Column 3 and 4 of table B27 present results with the dependent variable being equal to one if participant i perceives the fakes on average as *Very credible* or *Credible*. All coefficients decline substantially and are no longer statistically significant.[8] As we observe only small and not very robust differences for the order of the fact-check, we conclude that the influence of the order is most likeley not the driving force behind our main results. If the order of the fact-check played a role in our original experiment, the estimates we obtained there should

[7]A direct comparison between the two treatments indicates that participants of the MediaLiteracy group are 10.4 percentage points less likely to rate a post as *Very credible, Credible* or *Indecisive* ($p < 0.001$) than those of the FactChecking group

[8]The results are also not robust to any other specification of the cutoff.

probably be regarded as lower bounds, because the fact-checks were presented first in the original study. As we have emphasized earlier, we are cautious in interpreting the quantitative magnitude of a coefficient due to its dependency on the specific posts displayed, the time and the context in which the data was collected. Moreover, the qualitative nature of our findings with respect to the MediaLiteracy group remains unchanged. Even when fact-checks are displayed after the fake the effect for the FactChecking group is still substantially lower than for the MediaLiteracy group.

Table B27: Credibility of fakes by position of fact-checks – Fact-checking

	Default cutoff		**Stricter cutoff**	
	Environment			
	(1)	(2)	(3)	(4)
post	-0.069	-0.075	-0.013	-0.014
	[0.038]	[0.037]	[0.030]	[0.030]
p-value	(0.072)	(0.044)	(0.664)	(0.650)
mix	-0.046	-0.046	-0.007	-0.014
	[0.032]	[0.032]	[0.026]	[0.026]
p-value	(0.160)	(0.154)	(0.778)	(0.571)
Controls	no	yes	no	yes
N	1,340	1,340	1,340	1,340

Notes: Table B27 shows the OLS estimates of comparing three subgroups within the FactChecking group which received fake and fact-check in varying order. *post* corresponds to participants where the fact-check was displayed after the fake, *mix* corresponds to participants where it was displayed once before and once after, the baseline category are participants that saw the fact-check before the post. The dependent variable for column 1 and 2 is a dummy equal to one if participant i perceives the fakes on environment on average as *Very credible*, *Credible* or *Indecisive*. Column 3 and 4 show results for a stricter definition of the dependent variable that only takes the value one if participant i perceives the fakes on average as *Very credible* or *Credible*. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

B.5.3.5 High Credibility vs. Low Credibility Fakes

In our original experiment, participants received fakes on Covid-19 during the first part of the survey, which appeared to be relatively obvious. They have a significantly lower mean credibility compared to the nutrition fakes in the second part of the original survey in wave I and II. In the main study, we argue that the media literacy intervention, unlike fact-checking, is not tied to specific posts but instead raises the readers' general awareness. However, the effect of fact-checking on other, non-fact-checked fakes might depend on the credibility of the fact-checked post. To analyse whether fact-checking more credible fakes would enhance participants' critical evaluation of subsequent posts, we varied the exposure by presenting fakes previously rated with either a higher or lower credibility, as described in Appendix B.5.2.

Surprisingly, our manipulation did not work as intended: The credibility ratings for the fakes shifted significantly in the follow-up survey compared to the pre-study. Panel A of table B28 compares the average Likert scores of participants in the pre-study to those in the NoIntervention group of the follow-up for each environmental fake. In the pre-study, a fake post about the Statue of Liberty had an average credibility score of 3, and a fake about airplane chemical trails scored 2.8. In contrast, the lower-end fakes had mean scores of 2.3 and 2.2 (columns 3 and 4). These differences disappear across all fakes in the follow-up, with mean values ranging between 2.4 and 2.6 in the NoIntervention group. As described earlier, this suggests that the manipulation did not work as intended.

As the second part of the survey remained identical to the original experiment and the credibility of the fakes in the original experiment differs substantially from the credibility in the follow-up, we proceed with the second-best option to analyse the influence of credibility of the fake on the effectiveness of our interventions for subsequent posts: We compare the estimates for nutrition fakes (as reported above) in the follow-up to those of the original study (as reported in the main part of the paper).[9]

Panel B of table B28 displays the mean Likert scores of the NoIntervention group for fakes in the first part of the first wave of our original experiment. The credibility levels of fakes are more than one standard deviation lower compared to the follow-up experiment. Hence, we treat the follow-up as a case study for a fact-checking intervention, where the fakes have high credibility and compare these

[9] Note that these posts were identical in both experiments; however, differences in evaluation may arise due to changes in time and context. Comparing to the respective NoIntervention group should control for these differences, nevertheless, the exact size of the coefficients should be interpreted with caution.

findings to our original experiment, where the fact-checked fakes have low credibility.

Table B29 combines results from table 3.2 and table B26 in one table to facilitate the comparison. As discussed in Section B.5.3.3 *Reproducibility*, unlike our original study, this follow-up reveals a small but statistically significant effect for participants of the FactChecking group on non-fact-checked posts. This is in line with the hypothesis that fact-checks targeting less obvious fakes may lead readers to approach subsequent posts with greater criticism.

At the same time, these findings strongly support our interpretation that the media literacy intervention enhances skills that can be more broadly applied than fact-checking, as it appears significantly more effective, even in the context of the follow-up with more-credible fakes.

Table B28: Mean credibility per fake in Likert values

Panel A: Pre-Study vs. Follow-up

	Statue of Liberty	Chemtrails	ZDF-TV-Show	Mallorca
	(1)	(2)	(3)	(4)
Pre-study	3.000	2.800	2.308	2.200
	(0.686)	(0.789)	(0.630)	(0.422)
N	30	30	30	30
Follow-up	2.392	2.648	2.622	2.614
Std. Dev.	(1.157)	(1.226)	(1.151)	(1.159)
N	681	681	658	658

Panel B: Original experiment

	Sana clinic	Radioactivity
	(1)	(2)
Original experiment	1.979	1.652
Std. Dev.	(1.169)	(0.994)
N	618	618

Notes: Table B28 shows the mean credibility in Likert values from a 5-point Likert scale for all posts on the environmental topics for the participants of the pre-study compared to the NoIntervention group. The Likert scale ranges from 1 equals *very credible* to 5 equals *very incredible*. Standard deviation in parentheses.

Table B29: Immediate effects on credibility of fakes from a new topic by low cred. and high cred.

Panel A: Fact-checking				
	Low cred.		**High cred.**	
	(Original Experiment)		(Follow-up)	
	Nutrition			
	(1)	(2)	(3)	(4)
Fact-checking	-0.015	-0.010	-0.046	-0.046
	[0.018]	[0.017]	[0.015]	[0.015]
p-value	(0.385)	(0.553))	(0.002)	(0.002)
Controls	no	yes	no	yes
N	1,225	1,225	2,679	2,679
Panel B: Media literacy				
	Nutrition			
	(1)	(2)	(3)	(4)
Media literacy	-0.060	-0.061	-0.153	-0.148
	[0.019]	[0.019]	[0.016]	[0.016]
p-value	(0.002)	(0.001)	(0.000)	(0.000)
Controls	no	yes	no	yes
N	1,231	1,231	2,663	2,663
Baseline: No Intervention				
Mean DV	0.901	0.901	0.830	0.830
Std.Dev. DV	0.299	0.299	0.375	0.375

Notes: Table B29 shows the OLS estimates of comparing the NoIntervention to the FactChecking (Panel A) and to the MediaLiteracy group (Panel B), respectively. The dependent variable is a dummy equal to one if participant i perceives the **fakes** on nutrition on average as *Very credible*, *Credible* or *Indecisive*. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

B.5.3.6 Time spent with posts

In the main study we argue that the media literacy intervention raises participants' attention towards the Facebook postings. Our analysis shows that participants in the FactChecking group spend on average about 24 seconds less with the posts than those in the NoIntervention group, while participants in the MediaLiteracy intervention spend on average about 15 seconds more time with them. The estimates are highly statistically significant for both interventions. For the fact-checking intervention it corresponds to a reduction of about 25% of a standard deviation in the dependent variable in the baseline NoIntervention group and to about 16% of its mean value; for the media literacy intervention, it corresponds to an increase of about 16% and 10%, respectively.[10]

These findings are consistent with our proposed mechanism: While the media literacy intervention seems to increase the attention of participants, a fact-check seems to work like a heuristic which makes participants skip over posts faster. Details on the time spent with posts are presented in table B30.

B.5.3.7 Probability to fail an attention check

We do not find statistically significant differences in the probability to fail an attention check between our experimental groups. Details on these probabilities are denoted in table B31.

[10]The difference between the FactChecking and the MediaLiteracy group is statistically significant (two-sided t-test, $p = 0.000$).

Table B30: Time spent with fakes and facts

Panel A: Fact-checking		
	(1)	(2)
Fact-checking	-26.383	-23.447
	[3.457]	[3.290]
p-value	(0.000)	(0.000)
Controls	no	yes
N	2,679	2,679
Panel B: Media Literacy		
	(1)	(2)
Media Literacy	14.909	16.481
	[3.777]	[3.552]
p-value	(0.000)	(0.000)
Controls	no	yes
N	2,663	2,663
Baseline: No Intervention		
Mean DV	153.573	153.573
Std.Dev. DV	95.066	95.066

Notes: Table B30 compares the absolute time spent (in seconds) with the fakes and facts shown in the follow-up survey experiment for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B). All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Table B31: Probability to fail attention check

Panel A: Fact-checking		
	(1)	(2)
Fact-checking	0.001	0.002
	[0.016]	[0.016]
p-value	(0.971)	(0.912)
Controls	no	yes
N	2,679	2,679
Panel B: Media Literacy		
	(1)	(2)
Media Literacy	-0.016	-0.022
	[0.016]	[0.015]
p-value	(0.294)	(0.144)
Controls	no	yes
N	2,663	2,663
Baseline: No Intervention		
Mean DV	0.214	0.214
Std.Dev. DV	0.411	0.411

Notes: Table B31 compares the probability to fail any of our attention checks for participants from the NoIntervention to the FactChecking (Panel A) and the MediaLiteracy group (Panel B). All estimates are OLS estimates. Robust standard errors in squared parentheses, p-values in round parentheses. Control variables include age, gender, family status, household earnings, education, personality traits ("big 5"), political preferences, and prior knowledge on current events, health, and nutrition.

Appendix C

Appendix to Chapter 4

C.1 Omitted tables

C.1.1 Balance Table

Table C1: Balance Table

Variable	Category	Mean (Treated)	Mean (Control)	P-Value
Age		42.77	43.08	0.861
Male		0.568	0.637	0.257
Income	1500 € - 2500 €	0.328	0.363	0.556
Income	2500 € - 3500 €	0.248	0.281	0.543
Income	3500 € - 4500 €	0.096	0.111	0.691
Income	Not specified	0.016	0.015	0.938
Income	More than 4500 €	0.048	0.059	0.689
Income	Less than 1500 €	0.264	0.170	0.067*
Pol. Orientation	AfD	0.032	0.022	0.628
Pol. Orientation	CDU/CSU	0.096	0.081	0.682
Pol. Orientation	Die Grünen	0.392	0.400	0.896
Pol. Orientation	Die Linke	0.088	0.059	0.376
Pol. Orientation	FDP	0.040	0.059	0.478
Pol. Orientation	Not specified	0.024	0.044	0.370
Pol. Orientation	SPD	0.072	0.133	0.106
Pol. Orientation	Other	0.200	0.148	0.271
Pol. Orientation	I wouldn't vote	0.056	0.052	0.883

Notes: Age: age in years; Male: dummy equal to 1 if a person identifies as male and 0 otherwise; Income: disposable net income per month in Euros; Political orientation: Party a participant would vote for if there were national elections on the next day. The column *P-Value* reports p-values of a t-test for age and of a Chi2-test for male, income and political orientation. *** p<0.01, ** p<0.05, * p<0.1

C.1.2 Type Matches by Error Type

Table C2: OLS Estimates for Type Matches by Error Type

	Correlation VS. Causation		*Context*		*Interpretation of Statistics*	
	(1)	(2)	(3)	(4)	(5)	(6)
TG	0.102	0.092	0.119	0.057	0.230**	0.224**
	(0.068)	(0.068)	(0.098)	(0.108)	(0.093)	(0.100)
p-value (TG)	*0.133*	*0.176*	*0.226*	*0.601*	*0.016*	*0.027*
Age		0.005**		0.000		-0.003
		(0.002)		(0.003)		(0.003)
Male		-0.040		-0.078		-0.171*
		(0.073)		(0.109)		(0.100)
Earnings FE	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes
R^2	0.011	0.091	0.016	0.220	0.056	0.264
Observations	272	272	97	97	99	99

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the journalist in parentheses (260 clusters). p-values for TG are displayed at the bottom. Subsets refer to different article groupings. *** p<0.01, ** p<0.05, * p<0.1.

C.1.3 Main Results: IV Estimates

Table C3: Main Results: IV Estimates

	Accuracy of Written Headline		*Accuracy of Headline Classification*		*Abs. Distance to Classified Mistakes*		*Accuracy of Type*		*Z-score Index*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
treated	0.573***	0.589***	0.048	0.036	0.911***	0.889***	0.058	0.058	0.248**	0.257**
	(0.099)	(0.101)	(0.089)	(0.085)	(0.281)	(0.280)	(0.109)	(0.110)	(0.114)	(0.115)
p-value	[0.000]	[0.000]	[0.595]	[0.670]	[0.001]	[0.002]	[0.597]	[0.599]	[0.030]	[0.025]
Age		-0.001		0.001		-0.002		0.002		0.002
		(0.002)		(0.001)		(0.005)		(0.002)		(0.002)
Male		0.015		0.041		-0.376**		-0.033		0.043
		(0.052)		(0.046)		(0.151)		(0.060)		(0.060)
Earnings FE	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes
Constant	0.359***	0.212	0.500***	0.291	1.241***	0.936**	0.463***	0.453**	-0.060	-0.338
	(0.031)	(0.169)	(0.032)	(0.160)	(0.093)	(0.434)	(0.039)	(0.189)	(0.040)	(0.205)
R^2		0.037		0.067		0.038		0.042		0.034
Observations	520	520	520	520	520	520	520	520	520	520

Notes: This table reports 2SLS estimates with robust standard errors clustered at the level of the individual journalist in parentheses (260 clusters). A participant is regarded as treated if the person spend at least 5.6 minutes watching our video intervention. We use assignment to treatment as an instrumental variable. Each column corresponds to a separate regression model. *** p<0.01, ** p<0.05, * p<0.1.

C.1.4 Heterogeneity in Main Results

Table C4: OLS Estimates for Headline Correctness and Mistakes

	Accuracy of Written Headline		*Abs. Distance to Classified Mistakes*	
	(1)	(2)	(3)	(4)
TG	0.361***	0.357***	-0.019	-0.020
	(0.055)	(0.056)	(0.181)	(0.182)
High Earner	0.152**	0.145**		
	(0.062)	(0.065)		
Left			-0.205	-0.224
			(0.167)	(0.169)
TG × High Earner	-0.161*	-0.168*		
	(0.089)	(0.091)		
TG × Left			0.477*	0.456*
			(0.250)	(0.250)
Age		0.003		0.003
		(0.002)		(0.005)
Male		-0.037		-0.185
		(0.044)		(0.129)
Earnings FE / Vote FE	no	yes	no	yes
R^2	0.095	0.109	0.016	0.023
Observations	520	520	520	520

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the individual (260 clusters). *** p<0.01, ** p<0.05, * p<0.1.

C.1.5 Bonferroni Correction for Main Outcomes

Table C5: Main Results: OLS Estimates with Bonferroni Correction

	Accuracy of Written Headline		*Accuracy of Headline Classification*		*Abs. Distance to Classified Mistakes*		*Accuracy of Type*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TG	0.289	0.292	0.024	0.018	0.253	0.231	0.029	0.029
	(0.044)	(0.044)	(0.045)	(0.043)	(0.125)	(0.126)	(0.055)	(0.056)
Non-adjusted p-value	2.40×10^{-10}	1.97×10^{-10}	0.596	0.675	0.044	0.068	0.599	0.605
Bonferroni p-value	9.61×10^{-10}	7.89×10^{-10}	1.000	1.000	0.176	0.273	1.000	1.000
Controls	no	yes	no	yes	no	yes	no	yes
Constant	0.359***	0.120	0.500***	0.285	1.415***	1.870***	0.463***	0.444*
	(0.031)	(0.163)	(0.032)	(0.167)	(0.082)	(0.337)	(0.039)	(0.196)
R^2	0.083	0.112	0.001	0.069	0.008	0.029	0.001	0.041
Observations	520	520	520	520	520	520	520	520

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the individual journalist (260 clusters) in parentheses. Each column corresponds to a separate regression model. Both non-adjusted and Bonferroni-corrected p-values are provided for each hypothesis. *** p<0.01, ** p<0.05, * p<0.1.

C.1.6 Robustness Checks: Different Classifications for Headline Accuracy

Table C6: Different Classifications for Headline Accuracy

	Human Classifier 2		*ChatGPT*	
	(1)	(2)	(3)	(4)
TG	0.293***	0.296***	0.166***	0.165***
	(0.044)	(0.045)	(0.042)	(0.042)
p-value	[0.000]	[0.000]	[0.000]	[0.000]
Age		0.003		0.000
		(0.002)		(0.002)
Male		-0.034		-0.028
		(0.045)		(0.045)
Earnings & Politics FE	no	yes	no	yes
R^2	0.086	0.111	0.030	0.062
Observations	520	520	520	520

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the individual (260 clusters). *** p<0.01, ** p<0.05, * p<0.1.

C.1.7 Agreement of Classifications for different Iterations: Accuracy of Follow-Up Articles

Table C7: Pairwise Agreement

Pair of Iterations	**Cohen's κ**
Iteration 1 vs. Iteration 2	0.63
Iteration 1 vs. Iteration 3	0.83
Iteration 2 vs. Iteration 3	0.67

C.1.8 Followup: Analysis of Attrition

Table C8: Attrition Analysis: Consent and Publication of Science Articles

	Shared Name (1) Full Sample	*About Study* (2) Full Sample	*About Study* (3) Cond. on Consent
TG	0.143**	0.033	0.039
	(0.060)	(0.025)	(0.041)
Age	0.007***	0.000	-0.000
	(0.002)	(0.001)	(0.001)
Male	0.088	0.015	0.023
	(0.063)	(0.027)	(0.045)
Earnings FE	yes	yes	yes
Vote FE	yes	yes	yes
R^2	0.124	0.041	0.075
Observations	520	520	308

Notes: This table reports OLS estimates with robust standard errors clustered at the individual level (ID). Column (1) examines consent to share identifying information. Column (2) examines whether participants wrote about a scientific study. Column (3) restricts the sample to participants who consented to name sharing. Earnings and vote fixed effects are included in all specifications. *** p<0.01, ** p<0.05, * p<0.1.

C.1.9 Followup: Analysis of Demographics and Performance

Table C9: Demographics Check: Attrition Analysis

	Attrited (N=106)	**Follow-Up** (N=154)	**Difference** (Mean Diff.)	**p-value** (T-test)
Performance (Baseline)				
Headline Correctness (Score)	0.92	1.05	-0.14	0.156
Classified Correctly (Score)	1.11	0.96	0.15	0.099
Mistakes Distance (Abs.)	2.93	3.42	-0.48	0.133
Demographics				
Age (Years)	39.95	41.98	-5.03	0.013
Male (Share)	0.56	0.64	-0.08	0.196
High Income (Share)	0.06	0.05	0.01	0.228†

Note: The table compares demographic characteristics and performance of respondents who dropped out versus those who participated in the follow-up survey. P-values are based on two-sided t-tests (continuous) or Pearson χ^2 tests (categorical). † P-value for income refers to the distribution across all categories. *** p<0.01.

C.1.10 Followup Accuracy: IV Estimates per Iteration

Table C10: Followup Accuracy: IV Estimates per Iteration

	Main classification		*Iteration 1*		*Iteration 2*		*Iteration 3*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
treated	-0.444**	-0.378*	-0.276	-0.159	-0.192	-0.017	-0.556***	-0.443**
	(0.191)	(0.223)	(0.221)	(0.345)	(0.329)	(0.297)	(0.198)	(0.219)
p-value	[0.020]	[0.090]	[0.211]	[0.644]	[0.560]	[0.954]	[0.005]	[0.043]
Age		0.001		0.002		0.006		0.002
		(0.004)		(0.007)		(0.006)		(0.004)
Male		0.054		0.146		0.236***		0.063
		(0.099)		(0.163)		(0.078)		(0.099)
Earnings FE	no	yes	no	yes	no	yes	no	yes
Constant	0.444**	0.347	0.337	0.090	0.283	-0.129	0.556***	0.354
	(0.191)	(0.218)	(0.210)	(0.566)	(0.297)	(0.428)	(0.198)	(0.218)
R^2	0.165	0.221	0.105	0.153	0.136	0.348	0.270	0.350
Observations	58	58	58	58	58	58	58	58

Notes: Robust standard errors, which were bootstrapped in 21 clusters at the journalist level. "Treated" is instrumented with assignment to treatment (TG). Age (in years), Male (indicator), and Earnings fixed effects (four income brackets) are included where indicated. Stars denote significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

C.1.11 Different Dimensions of Headline Classifications

Table C11: Definitions of constructed headline variables

Variable	Definition and coding rule
sentiment	Polarity indicator. Takes the values -1 (negative tone), 0 (neutral) or 1 (positive).
emotionality	Continuous score in the closed interval $[0, 1]$ capturing the share of *affective* relative to *cognitive* language, following the emotion–reason dimension proposed by Gennaro and Ash (2022). Higher values indicate more emotional wording.
specificity	Continuous index in $[0, 1]$ that measures how narrowly the headline's subject matter is defined. A value of 0 corresponds to extremely broad headlines (e.g. "Markets rally"), whereas 1 denotes highly detailed references to actors, magnitudes or events.
newsworthy	Continuous index in $[0, 1]$ reflecting the ex-ante news value of the event, based on the proper-scoring-rule concept of newsworthiness developed by Armona et al. (2024). Larger values indicate headlines that are both surprising and consequential to a general audience.
clickbait	Continuous index in $[0, 1]$ that quantifies the degree of curiosity-gap or sensational framing typical of "clickbait". Values near 1 denote strong clickbait cues; values near 0 correspond to purely factual, literal headlines.
density	Continuous index in $[0, 1]$ that proxies the amount of factual information per token. Higher scores correspond to shorter yet detail-rich headlines; lower scores reflect verbose or filler language.
extreme	Binary indicator equal to 1 if the headline employs extremist, inflammatory, or all-caps hyperbole (e.g. "DISASTER", "traitors"), and 0 otherwise.

C.1.12 Results: Linguistic Headline Classifications

Table C12: Results: Treatment Effects on Linguistic Features of Headlines

	Sentiment		*Emotionality*		*Specificity*		*Newsworthiness*		*Clickbait*		*Density*		*Extreme*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
TG	0.109**	0.117**	-0.063***	-0.061***	0.005	0.003	-0.009	-0.009	-0.011	-0.008	0.009	0.007	0.013	0.010
	(0.054)	(0.055)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.017)	(0.017)	(0.020)	(0.019)	(0.013)	(0.013)
p-value	[0.044]	[0.034]	[0.000]	[0.001]	[0.790]	[0.866]	[0.633]	[0.630]	[0.502]	[0.637]	[0.637]	[0.708]	[0.298]	[0.439]
Age		-0.005**		0.000		-0.001		0.000		-0.000		-0.001		-0.000
		(0.002)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)		(0.001)
Male		0.028		0.011		-0.016		-0.007		0.023		-0.014		0.034***
		(0.055)		(0.018)		(0.019)		(0.020)		(0.017)		(0.020)		(0.012)
Income FE	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Politics FE	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
Constant	-0.295***	-0.262	0.393***	0.465***	0.568***	0.459***	0.538***	0.459***	0.353***	0.445***	0.590***	0.508***	0.015**	-0.026
	(0.040)	(0.226)	(0.013)	(0.088)	(0.014)	(0.093)	(0.014)	(0.092)	(0.013)	(0.077)	(0.015)	(0.080)	(0.007)	(0.027)
R^2	0.007	0.031	0.026	0.072	0.000	0.056	0.000	0.022	0.001	0.037	0.001	0.051	0.002	0.044
Observations	516	516	516	516	516	516	516	516	516	516	516	516	516	516

Notes: Robust standard errors clustered at the journalist level (258 clusters) are shown in parentheses. *sentiment* is a polarity indicator taking values -1 (negative), 0 (neutral) or 1 (positive); *emotionality* is a continuous score in $[0, 1]$ measuring the share of affective versus cognitive language; *specificity* is an index in $[0, 1]$ that rises with the concreteness of actors, magnitudes or events named in the headline; *newsworthy* is a $[0, 1]$ index based on the proper-scoring-rule concept of news value in Armona et al. (2024); *clickbait* is a $[0, 1]$ index capturing curiosity-gap or sensational framing cues; *density* is a $[0, 1]$ proxy for factual content per token (higher values denote concise yet information-rich headlines); *extreme* is a binary indicator equal to 1 if the wording uses extremist, inflammatory, or all-caps hyperbole and 0 otherwise. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C.1.13 Results: Linguistic Headline Cosine Similiarity

Table C13: Two-Sample t-test for Cosine Similarity

Group	Obs.	Mean	Std. Error	Std. Dev.	95% Conf. Interval
Untreated (0)	285	0.113	0.010	0.175	[0.093, 0.133]
Treated (1)	263	0.104	0.011	0.179	[0.082, 0.126]
Combined	548	0.109	0.008	0.177	[0.094, 0.123]
Difference		0.009	0.015		[-0.021, 0.039]

Notes: This table reports the results of a two-sample t-test with equal variances. The difference is calculated as Mean(Untreated) - Mean(Treated).
The null hypothesis is H0: diff = 0. The two-sided alternative (Ha: diff $\neq$ 0) yields a t-statistic of 0.586 with 546 degrees of freedom. The corresponding p-value is 0.559.
*** p<0.01, ** p<0.05, * p<0.1

C.1.14 Heterogeneity: Continuous Age Control

Table C14: Heterogeneity Analysis: Continuous Age Interaction

	Accuracy of Written Headline	*Accuracy of Headline Classification*	*Abs. Distance to Classified Mistakes*	*Accuracy of Type*	*Z-score Index*
	(1)	(2)	(3)	(4)	(5)
Treated	0.819***	0.074	-0.289	0.005	0.179
	(0.238)	(0.135)	(0.469)	(0.143)	(0.891)
Age	0.008*	0.002	-0.007	-0.001	0.009
	(0.004)	(0.002)	(0.007)	(0.002)	(0.014)
Treated × Age	-0.007	-0.001	0.015	0.002	-0.011
	(0.005)	(0.003)	(0.011)	(0.003)	(0.019)
Male	Yes	Yes	Yes	Yes	Yes
Earnings FE	Yes	Yes	Yes	Yes	Yes
Politics FE	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes
R^2	0.200	0.069	0.033	0.057	0.078
Observations	520	520	520	520	520

Notes: This table reports OLS estimates with robust standard errors clustered at the level of the individual journalist in parentheses. The dependent variables correspond to those in the main results. *Age* is measured as a continuous variable. All models include the full set of covariates used in the main analysis: gender, earnings categories, and political party preference. *** p<0.01, ** p<0.05, * p<0.1.

C.1.15 Heterogeneity: Political Leaning

Table C15: Heterogeneity Analysis: Political Orientation (Multi-Category)

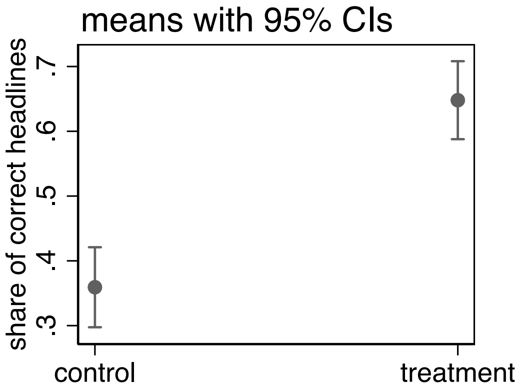
	Written accuracy	Headline classif.	Abs. dist. mistakes	Type accuracy	Index (z-score)
	(1)	(2)	(3)	(4)	(5)
Treated (Base Group)	0.322	0.614	-0.024	-0.157	-2.003
	(0.211)	(0.488)	(0.516)	(0.278)	(1.455)
Interactions (Difference to Base Group)					
Treated × Party 2	-0.309	0.127	0.462	0.349	1.201
	(0.278)	(0.546)	(0.770)	(0.319)	(1.710)
Treated × Party 3	-0.218	-0.322	0.670	0.164	1.554
	(0.221)	(0.501)	(0.654)	(0.286)	(1.552)
Treated × Party 4	-0.180	0.286	0.425	0.211	0.478
	(0.260)	(0.527)	(0.694)	(0.317)	(1.790)
Treated × Party 5	0.016	0.186	0.626	0.321	1.872
	(0.339)	(0.595)	(0.806)	(0.364)	(2.034)
Treated × Party 6	-0.392	0.445	-0.122	-0.232	1.400
	(0.280)	(0.623)	(0.902)	(0.359)	(1.985)
Treated × Party 7	-0.203	0.060	0.669	0.131	2.484
	(0.276)	(0.537)	(0.670)	(0.295)	(1.705)
Treated × Party 8	-0.307	-0.110	-0.348	0.173	2.696*
	(0.239)	(0.523)	(0.603)	(0.295)	(1.570)
Treated × Party 9	-0.111	-0.441	-0.109	0.224	1.666
	(0.273)	(0.543)	(0.673)	(0.317)	(1.877)
Male	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes
Earnings FE	Yes	Yes	Yes	Yes	Yes
Observations	520	520	520	520	520
R^2	0.064	0.228	0.043	0.078	0.098

Notes: This table reports OLS estimates testing for treatment effect heterogeneity across political groups. The reference category is Party 1 (likely CDU/CSU). The interaction terms indicate whether the treatment effect differs from the base group. All models include controls for age (continuous), gender, and earnings fixed effects. Robust standard errors clustered at the journalist level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

C.2 Omitted figures

C.2.1 Results

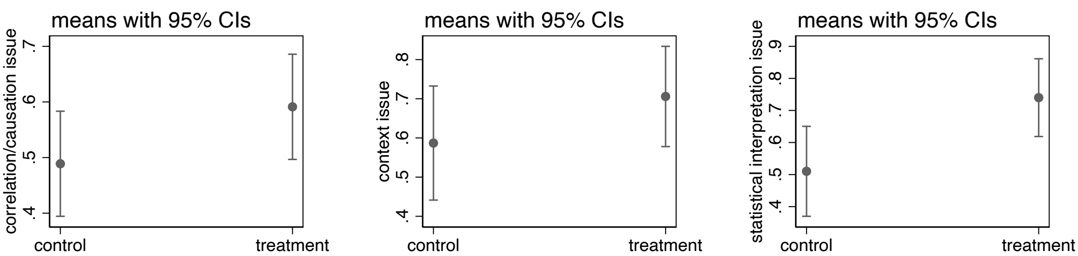
Figure C1: Main Result: Improvements in Headline Accuracy



Notes: Figure C1 depicts the means and 95-percent confidence intervals of the share of headlines classified as factually accurate. The estimates displayed here are obtained by running the baseline regression as described in equation C.1 (without controls) with robust standard errors clustered on the level of the journalist (260 clusters).

Figure C2: Results: Identification of Error Types by Type of Error

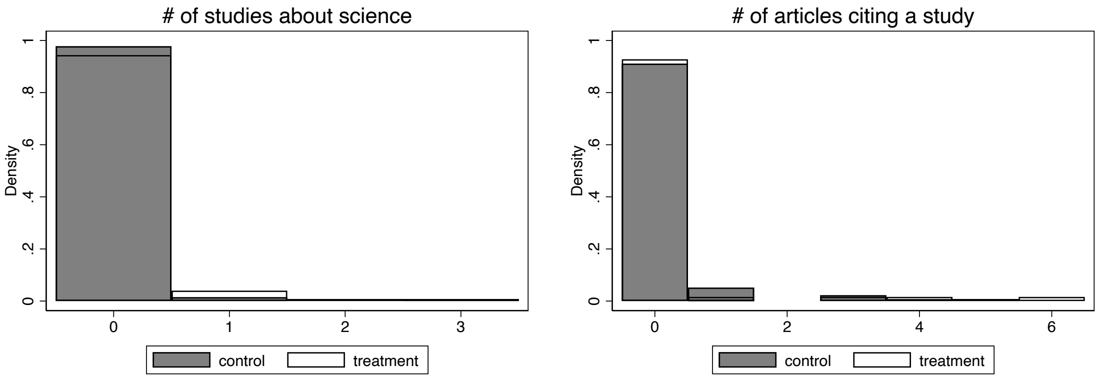
(a) Correlation VS. Causation **(b)** Context **(c)** Interpretation of Statistics



Notes: Figure (a) illustrates the means and 95-percent confidence intervals for the correct classification of the error type for articles with issues in the area correlation VS. causation, (b) for articles with issues in the explanation of context and (c) with issues in the interpretation of statistics. The estimates displayed here are obtained by running the baseline regression as described in equation C.1 (no controls).

Figure C3: Results: Science reporting after participation

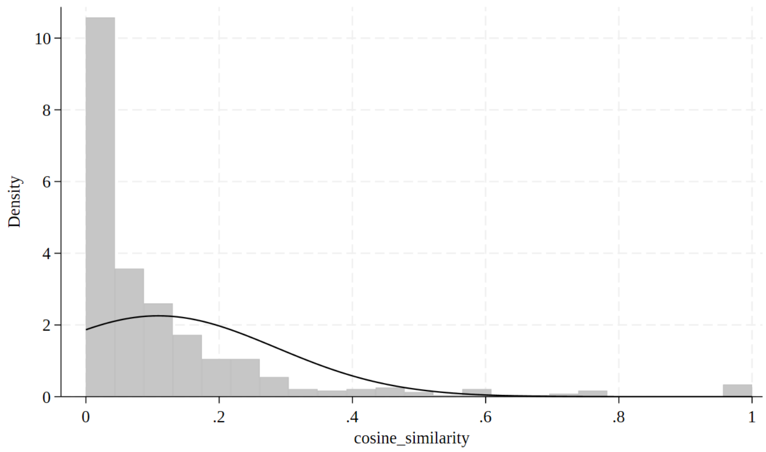
(a) Distribution **(b)** Distribution



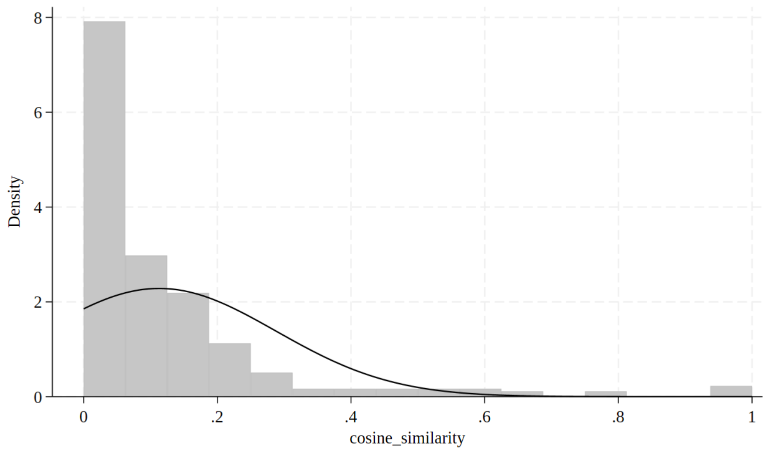
Notes: Figure (a) illustrates the distribution of the number of studies that have a scientific article as their main topic. Figure (b) depicts the distribution of articles citing at least one study. The numbers always refer to how many out of the 10 last articles published by a journalist after participation in our study fulfill the criterion. Values for the control group are shaded in gray.

Figure C4: Distribution of Cosine Similarity Scores

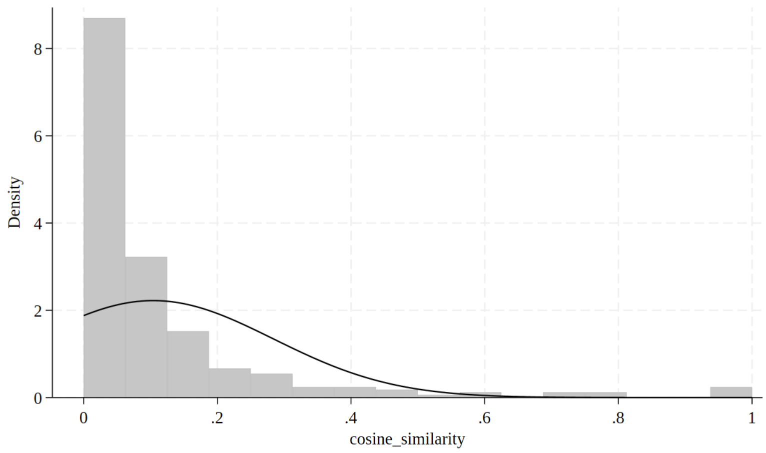
(a) Full Sample



(b) Control Group



(c) Treatment Group



Notes: The figure shows distributions of cosine similarity between original and rewritten headlines. Panel (a) shows the full sample, panel (b) the control group, panel (c) the treatment group. A value of 0 corresponds to completely distinct headlines.

C.3 Pre-Study with Journalism School

We collaborated with the *Cologne School of Journalism* (*Kölner Journalistenschule*) for a pre-study in July 2022. We designed a 4-hours-long seminar with the goal to enhance the journalism students' science literacy and critical interpretation skills. We based the content of this seminar on the concepts which have been found effective in the literature on statistical and science education. The literature outlines diverse frameworks and concepts as essential to foster science literacy. Gal (2002) for instance defines it as a multifaceted skill encompassing mathematical knowledge, critical evaluation, and contextual understanding. Sharma (2017) highlight the importance of integrating cognitive and dispositional elements, such as critical thinking and reasoning, into training modules. Rumsey (2002) advocate for a "statistical citizenship", which emphasizes reasoning and decision-making in everyday contexts. Schield (1999) underscore the need to differentiate between correlation and causation, sample versus population, and the interpretation of statistical tests. These conceptualizations collectively emphasize that science literacy goes beyond computational skills, requiring critical engagement with data and its societal implications. We incorporated the key concepts outlined above into a seminar and evaluated its effectiveness using a pre-post design based on participants' performance in short quizzes. The seminar comprised two components. The first component introduced students to foundational concepts in statistics and scientific methodology. Participants engaged in applied exercises and attended a lecture segment focused on interpreting the results sections of scientific studies. This segment addressed statistical inference, conditional probabilities, appropriate interpretation of effect sizes, and issues related to sampling. The second component emphasized practical strategies for assessing the credibility of scientific research from a journalistic standpoint. Topics included the use of academic search engines, effective questioning in interviews with researchers, and the evaluation of research funding sources. Participants learned to distinguish between high-quality research capable of supporting causal inference and studies based on weaker designs. These skills were reinforced through applied exercises.[1] In the first component, the average share of correct responses in quiz questions increased from 65 percent before the seminar to 84 percent after participation. In the second component, the share rose from 43 percent to 73 percent. We also collected qualitative feedback from participants and drew on both this feedback and the quiz domains showing the largest room for improvements to inform the development of an educational video on science reporting for journalists.

[1]All seminar materials are available in our OSF repository.

C.4 Specifications

C.4.1 Baseline Estimation

Our baseline estimation is described by the following equation:

$$Y_i = \beta_0 + \beta_1 \mathrm{TG}_i + \beta_2 \mathbf{X}_i + \varepsilon_i \tag{C.1}$$

where:

- Y_i represents the outcome variable of interest for article i,
- TG_i is a treatment group indicator,
- $\mathbf{X}_i$ is a vector of covariates (age, gender, income, political orientation), and
- ε_i is the error term.

Note that the regression is conducted at the article level, yielding two observations per participant, as each journalist evaluated two studies with a corresponding article. To account for within-participant correlation, we cluster standard errors at the journalist level, resulting in 260 clusters.

C.4.2 IV Estimation (Robustness)

In our setup, it is possible that some participants in the treatment group bypassed the intervention by quickly clicking through the survey pages without watching the full video. In such cases, the baseline analysis presented in Section C.4.1 may underestimate the true treatment effect. To account for this, we define a binary variable $time_i$, which equals 1 if participant i spent a reasonable amount of time engaging with the video. Specifically, we set $time_i = 1$ if participant i spent more time on the video page than the median treatment group participant (5.6 minutes). We estimate the following two-stage least squares (2SLS) specification:

$$Y_i = \gamma_0 + \gamma_1 \widehat{time_i} + \gamma_2 X_i + \epsilon_i, \tag{2nd stage}$$

$$time_i = \pi_0 + \pi_1 TG_i + \pi_2 X_i + u_i, \tag{1st stage}$$

As before, standard errors are clustered at the level of the individual journalist (260 clusters).

C.4.3 IV Estimation (Follow-Up)

Our follow-up data (described in Section 4.3.1.4) is structured at the level of individual published articles. A key limitation of this data is the presence of potential "always takers" in the control group. These are participants who accessed the educational video after completing the survey. Specifically, 37 percent of control group participants clicked on the video link post-participation. To address the resulting endogeneity, we estimate a 2SLS-model in which participants are coded as treated if they either belonged to the treatment group or subsequently accessed the video. Assignment to the treatment group serves as the instrumental variable. The model is specified as follows:

$$Y_i = \gamma_0 + \gamma_1 \widehat{treated}_i + \gamma_2 X_i + \epsilon_i, \qquad \text{(2nd stage)}$$

$$treated_i = \pi_0 + \pi_1 TG_i + \pi_2 X_i + u_i, \qquad \text{(1st stage)}$$

For the analysis of whether an article is about science or cites a scientific study, we cluster standard errors at the journalist level (64 clusters). For the analysis of factual accuracy in the subsample of articles that either focus on science or cite a study, we bootstrap standard errors at the journalist level, given the smaller sample size (58 articles authored by 21 journalists), which is insufficient for reliable standard clustering methods.

C.5 Script of the educational video (translation to English)

As our participants are German journalists, we provided them with an educational video in German. It can be accessed here: `https://youtu.be/0R322UjAY9Q`. *A translation to English of the videos' audio content is provided below. Any content that was not part of the audio/text (e.g. as illustrations) is described in blue, italic text.*

Welcome to this brief introduction to interpreting statistics.

Numbers and statistics are everywhere. As journalists, you likely need to read studies and summarize their results concisely. But interpreting statistics correctly isn't always straightforward. In this video, we want to share a few simple rules with you.

First, we need to ask ourselves three questions:

1. Who commissioned the study?
2. How did the researchers conduct their work?
3. How are the results presented?

The three points appear next to the speaker as bullet points; Point 1 is enlarged as a headline. Visuals: Wallet and money being exchanged.

Who commissioned the study?

Whoever commissions or funds a study may influence its outcomes. Pay close attention to whether a study is independent research or commissioned work—such as for companies or interest groups—and interpret the results cautiously.

The headline changes; Point 2: "How did the researchers conduct their work?"

How did the researchers conduct their work?

The term "study" isn't protected. Studies can vary greatly in quality. However, with simple guidelines, you can quickly assess how reliable the results are.

Points appear as bullet points.

Topic: Sample. Who or what is being studied? Who does this sample represent? For example, the entire population or just a specific group?

The selection of the sample is crucial. If only a specific population group is covered, the results cannot easily be generalized to others. For example, if only DAX board members are surveyed about their household income, you can't infer the median income of the general population from that.

Samples can also be too small to draw general conclusions. There's no universal rule for an appropriate sample size. However, you can ask yourself: How large is the sample relative to the whole group? Does the sample represent the group in all characteristics (e.g., gender, age, location)? Use your common sense!

A graphic of a diverse group of people vs. only men appears. The headline changes; Point 3: "How are the results presented?"

How are the results presented? This is critical! Consider these four questions:

1. What metrics are used, and what don't they tell me?
2. Are absolute or relative frequencies mentioned?
3. Is it correlation or causation?
4. What do the graphs suggest, if used?

Points appear as bullet points; Point 1 is emphasized; Visualization optional.

Topic: "Metrics"

All metrics, such as averages and medians, convey certain information. At the same time, they omit a lot. For example, the average household income tells you nothing about income distribution—how far apart the highest and lowest incomes are. Different income distributions can result in the same average.

Visualization: Income distribution on a number line. Different distributions lead to the same average.

Topic: “Frequencies”

Absolute and relative frequencies are often misinterpreted. Absolute frequency indicates how often an event occurs. For example, if 1 in 100,000 people contracts a rare disease, the absolute frequency is 1. The relative frequency is 1/100,000, or 0.001 percent. If 2 out of 1,000 people contract the same disease, the relative frequency increases by 100 percent! Yet, it still only represents 0.002 percent, meaning only 2 out of 100,000 people are affected. Percent increases in small absolute and relative probabilities can be misleading. It’s often worth examining them alongside absolute probabilities.

Calculation appears as a visualization next to Anna, possibly with highlighted boxes.

Topic: “Correlation and Causation”

Be cautious, especially with correlations and causations. Does one phenomenon merely occur alongside another? Or does one cause the other? A well-known example:

Text appears: “Playing violent video games makes teens aggressive and leads to school shootings.”

Media and politicians have repeatedly discussed this topic. However, scientists agree: There’s only a correlation here. Potentially aggressive teens are often the ones who enjoy playing violent video games. But that doesn’t mean the games cause aggression. The two phenomena simply occur simultaneously without a causal link.

Finally, let’s look at graphs:

Two graphs appear showing the development of a fund; one starts the Y-axis at zero, the other slightly below the paper value in the first year.

Keep in mind: Graphs often carry a pre-determined interpretation. Here, we see the development of an equity fund. However, the graphs differ despite showing the same data: Depending on where the Y-axis begins, the positive growth of the fund appears much stronger.

When manipulating graphs, there are no limits to creativity: Misleading color choices, suggestive symbols like oversized arrows or skulls, and incorrect scales are all possible.

The graphs fade out.

To sum up briefly:

Reading and interpreting scientific studies can be challenging. Our tips: Follow these basic rules:

1. Who funded the study?
2. Is the sample representative?
3. What do the study's metrics say—and what don't they say?
4. Are absolute or relative frequencies used?
5. Correlation or causation?
6. Critically examine the graphs.

We hope this brief introduction was helpful. If you have further questions, feel free to contact us directly at the ifo Institute. We'd be happy to assist.

Contact information appears.

C.6 Science-related articles (translation to English)

This Appendix provides translations of the science-related articles participants were exposed to to English. Links to the studies the articles report on as well as the original articles in the German media are included in blue, italic text below the translations.

Rainforest 2

Surprising Study: Deforestation Promotes the Outbreak of Infectious Diseases

French researchers have found a significant statistical link between deforestation and outbreaks of infectious diseases transmitted by animals. A similar trend was observed with palm oil plantations: the larger their area, the more frequent the occurrence of infectious diseases. Another finding of the study, which was recently published in the journal "Frontiers in Veterinary Science," is that reforestation also leads to more cases of such diseases.

...

They found a strong connection between deforestation and epidemics, such as malaria and Ebola, in tropical countries like Brazil, Peru, Bolivia, the Democratic Republic of Congo, Cameroon, Indonesia, Myanmar, and Malaysia. In contrast, temperate regions like the USA, China, and Europe showed clear links between reforestation and diseases such as Lyme disease, which is transmitted by ticks.

...

Original study: *https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2021.661063/full.*
Article: *https://weather.com/de-DE/wissen/klima/news/2021-03-24-mehr-infektionen-was-abholzung-von-waldern-mit-unserer-gesundheit-tut.*

Rainforest 3

STUDY SHOWS: IF WE CONTINUE DEFORESTATION, THERE WILL BE MORE INFECTIOUS DISEASES

At first glance, deforested or reforested areas and infectious diseases might not seem to have much in common. However, scientists have now taken a closer look at the numbers.

French researchers have found a significant statistical link between deforestation and outbreaks of infectious diseases transmitted by animals. A similar trend was observed with palm oil plantations: the larger their area, the more frequent the occurrence of infectious diseases. Another finding of the study, which was recently published in the journal "Frontiers in Veterinary Science," is that reforestation also leads to more cases of such diseases.

...

They found a strong connection between deforestation and epidemics, such as malaria and Ebola, in tropical countries like Brazil, Peru, Bolivia, the Democratic Republic of Congo, Cameroon, Indonesia, Myanmar, and Malaysia. In contrast, temperate regions like the USA, China, and Europe showed clear links between reforestation and diseases such as Lyme disease, which is transmitted by ticks.

...

Original study: https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2021.661063/full. *Article:* https://www.dropbox.com/scl/fi/8xpg9yqv06uk7srgsv18k/rainforest_article3.pdf?rlkey=yhktzma835sitfoceajdppjm4&dl=0.

Vegetarian 1

Study: Vegans Are Significantly Less Likely to Suffer Severe COVID-19

U.S. researchers have found that diet can influence the course of COVID-19: According to the results, vegetarians and vegans have a 73% lower chance of developing moderate to severe COVID-19 compared to meat eaters. Those who also eat fish have a 59% lower risk than those who eat meat.

...

The researchers analyzed data from study participants in six countries, including Germany. They considered factors such as weight, age, pre-existing conditions, and lifestyle of the participants — according to the researchers, these factors did not have an influence on the results.

Original study: https://nutrition.bmj.com/content/early/2021/05/18/bmjnph-2021-000272. *Article:* https://www.nudefood.de/blog/veganerinnen-erkranken-deutlich-seltener-schwer-an-covid-19/.

Vegetarian 2

Diet Determines COVID Course: Are Vegetarians Less Likely to Get Seriously Ill?

Avoiding meat not only helps the environment but also benefits one's health — even in the case of a coronavirus infection. U.S. researchers have found that vegetarians and vegans generally fare better in a COVID-19 infection compared to meat eaters.

There are many factors that can influence a COVID-19 infection: gender, age, blood type. People with obesity and type 2 diabetes — both of which can result from an unhealthy lifestyle characterized by a lack of exercise and poor diet — are also at higher risk of severe illness. A U.S. research team has now taken a closer look at how a person's diet can affect the course of a COVID-19 infection.

...

The result: People who are vegetarians or vegans, meaning they eat only plant-based foods, or who also eat fish like pescatarians, were significantly less likely to suffer severe COVID-19 compared to meat eaters. According to the study, vegetarians and vegans had a 73% lower chance of developing moderate to severe COVID-19 compared to those who did not follow a strictly plant-based diet. For pescatarians, the risk was 59% lower compared to meat eaters.

Original study: https://nutrition.bmj.com/content/early/2021/05/18/bmjnph-2021-000272. *Article:* https://www.n-tv.de/wissen/Erkranken-Vegetarier-seltener-schwer-article22610491.html.

Vegetarian 3

Study Provides Evidence: This Diet May Increase COVID Risk

There are many factors that can influence a COVID-19 infection: gender, age, pre-existing conditions, blood type — and also our diet? Researchers claim to have found new evidence for this.

...

The result after analyzing the data: People who follow a plant-based diet — vegetarians and vegans — or consume only fish, like pescatarians, were significantly less likely to suffer severe COVID-19 compared to meat eaters.

Specifically: Vegetarians and vegans had a 73% lower chance of developing moderate to severe COVID-19 compared to those who did not follow a plant-based diet. For pescatarians, the risk was 59% lower compared to people who followed a different diet.

These trends remained consistent even after the researchers took into account participants' weight, age, pre-existing conditions, and lifestyle factors (such as smoking and physical activity). "In six countries, a plant-based or pescatarian diet was associated with a lower likelihood of moderate to severe COVID-19 illness," the study concluded. Why these results occurred remains unclear.

...

Original study: https://nutrition.bmj.com/content/early/2021/05/18/bmjnph-2021-000272. *Article:* https://www.t-online.de/gesundheit/krankheiten-symptome/id_90178816/corona-erkranken-fleischesser-schwerer-an-covid-19-studie.html.

Skoda 1

Scrap Car: From Car Statistics to Driver IQ

The first "Unstatistic of the Month" on marktforschung.de is the current study by the British comparison portal "Scrap Car." According to the widely published results, the choice of car can indicate the intelligence of its drivers. A bold claim that the Unstatistic team decided to examine more closely.

The "Unstatistic of the Month" is a study by the British comparison portal "Scrap Car" for used cars, suggesting that the car brand, the type of engine, or even its color can reveal a lot about the driver's intelligence. This was reported by sources such as sueddeutsche.de, autobild.de, and t-online.de. According to the results of these studies, Skoda drivers have the highest IQ, followed by owners of Suzuki and Peugeot vehicles. The "dumbest" drivers are those of BMW, Fiat, and Land Rover.

The color and type of engine of a car are also said to indicate driver intelligence: white and gray as well as gasoline-powered cars supposedly have relatively smart drivers. In contrast, if a car is silver or green, or an electric vehicle, the driver is supposedly less intelligent. And if you personalize your license plate, you are also more likely to be "dumb".

...

However, these values are only accurate for an (infinitely) large population, not for a small sample of the population. And this is precisely where the main problem of the above claims lies. When using samples, the estimated averages (in this case, IQ scores) can randomly deviate from the (true) average in the population.

...

Original study: https://www.scrapcarcomparison.co.uk/blog/smartest-drivers/. *Article:* https://www.marktforschung.de/marktforschung/a/scrap-car-studie-vom-auto-zum-iq/.

Skoda 2

2,000 Participants: Correlation Between Car Brand and Driver IQ Discovered

A British study shows that there are connections between car brands and driver IQ. Drivers of gasoline-powered cars are said to have the highest IQ, while drivers of electric cars have the lowest.
A study conducted by the market research institute Censuswide and the portal ScrapCarComparison discovered a correlation between the intelligence quotient (IQ) of drivers and the brand of their cars. About 2,000 people participated in the study, first taking an IQ test and then indicating which car they owned. The researchers also asked about the engine type, color, and license plate personalization of the vehicles.
According to the study data, people who own a green Land Rover with a personalized license plate are the drivers with the lowest intelligence. On average, Land Rover owners have an IQ of 88.58. Skoda drivers have the highest IQ, with an average of 99. Among German manufacturers, Mercedes drivers have the highest IQ, averaging 94.74. In the overall ranking, Mercedes is in 8th place.

...

In the research community, there are already doubts about the validity of the study, as the number of participants is relatively small. Therefore, the findings suggesting that one can infer a person's intelligence based on their car should be taken with a grain of salt.

...

Original study: `https://www.scrapcarcomparison.co.uk/blog/smartest-drivers/`. *Article:* `https://shorturl.at/HSWkt`.

Skoda 3

What a Car Brand Says About the Driver's Intelligence

What do car brand, color, and engine type say about the intelligence of the driver? Quite a lot, emphasizes a British study.

Skoda — Simply Clever? The slogan of the Czech car manufacturer seems to be true. At least, this is confirmed by a British study conducted by the comparison portal Scrap Car Comparison in collaboration with the market research institute Censuswide.

The portal investigated how IQ correlates with car brand — essentially, what one's car says about one's intelligence.

...

The result: (Of all things) Land Rover drivers are at the bottom with an average IQ of 88.58. At the top are Skoda drivers, with an average IQ of 99. They are followed by Suzuki (98.09) and Peugeot (97.79).

For comparison: Most people, around two-thirds, have an IQ between 85 and 115.

...

Original study: https://www.scrapcarcomparison.co.uk/blog/smartest-drivers/. *Article:* https://www.autobild.de/artikel/auto-marke-iq-intelligenz-motor-antrieb-e-auto-22596383.html.

Climate 1

Heat Increases the Risk of Premature Birth

Due to global warming, summers are becoming hotter in Germany as well. According to a study, this significantly increases the risk of late premature births.

It is well known that climate change will lead to longer and hotter summers in the Northern Hemisphere in the future. According to a new study by the University Medical Center Hamburg-Eppendorf (UKE), this could also significantly increase the risk of late premature births. Starting at 30 degrees Celsius, the risk of early delivery already increases by 20%. At temperatures above 35 degrees Celsius, the risk rises by up to 45%.

...

By 2033, considering current climate projections, almost one in six children could be born prematurely due to rising temperatures. That would be twice as many as today. "The consequences for the health of newborns are not yet foreseeable," Arck explained. The study's conclusion emphasizes the importance of implementing specific measures in healthcare now. Professionals, such as obstetricians, should not only be warned of such risks but guidelines for actively monitoring pregnant women during heatwaves should also be developed.

...

Original study: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00216-5/fulltext. *Article:* https://shorturl.at/UPbyj.

Climate 2

Study by Hamburg University Hospital: Climate Change Leads to More Premature Births

A study by Hamburg University Hospital is the first to establish a link between high temperatures and premature births. Heat stress can significantly increase the likelihood of premature birth.

...

Exactly this group of premature babies — currently around 70,000 children per year in Germany — is affected by climate change. If temperatures rise above 35 degrees Celsius, according to the UKE analysis, the risk of premature birth increases by 45%.

For the first time, researchers were able to establish a direct link between rising temperatures and direct impacts on human life. According to the study, heat stress caused by temperatures of 30 degrees leads to a 20% increase in premature births, and above 35 degrees, almost every second pregnant woman is affected.

The study's forecast suggests that in ten years, if hot summer days continue to become more frequent, every sixth child in Germany will be born prematurely — twice as many premature births as today.

...

Original study: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00216-5/fulltext. *Article:* https://www.zdf.de/nachrichten/panorama/fruehgeburten-hitze-anstieg-klimawandel-studie-100.html.

Climate 3

Heat Stress Increases the Risk of Spontaneous Premature Births

Sweltering summer heat is a particular burden for women in late pregnancy. Researchers in Hamburg have now found that the risk of premature birth increases on hot days.

Prolonged summer heat stress increases the risk of premature birth for women in late pregnancy, according to a new study. Several consecutive days with temperatures of 30 degrees Celsius increased the risk for women in the 34th to 37th week of pregnancy by 20%, and several days above 35 degrees increased the risk by 45%, according to a study led by researchers from the University Medical Center Hamburg-Eppendorf (UKE).

The team led by Professors Petra Arck and Anke Diemert analyzed more than 42,000 patient records of women who gave birth at the Hamburg hospital over the past 20 years.

The researchers noted in the journal "eBioMedicine" that expectant mothers can apparently handle one or two hot days well. It is only on the third, fourth, or fifth day without a break from the heat that premature labor becomes more frequent, especially when high humidity further increases the perceived heat.

...

Original study: https://www.thelancet.com/journals/ebiom/a *rticle/PIIS2352-3964(23)00216-5/fulltext. Article: https://www.fr.de/wissen/hitzestress-vergroessert-risiko-von-spontanen-fruehgeburten-zr-92361227.html.*

Weight 1

Study Shows Weight Gain During the COVID-19 Pandemic: Study Author Hans Hauner in Interview

According to a new study by TU Munich, Germans gained an average of 5.6 kilograms during the COVID-19 pandemic. The main reasons are emotional stress and lack of physical activity. The greatest weight gain was observed among those aged 30 to 44.

Almost every second German has gained weight since the start of the COVID-19 pandemic, averaging 5.6 kilograms. This is the finding of a new study by the Else Kröner Fresenius Center for Nutritional Medicine (EKFZ) at the Technical University of Munich (TUM).

Study author Professor Hans Hauner, Director of the EKFZ and Professor of Nutritional Medicine at TUM, shared his tips for balanced "pandemic nutrition" and his tricks for tackling excess "pandemic pounds" in an interview with the Sonntagsblatt.

...

Original study: https://www.dropbox.com/scl/fi/rwyenyzslphws5gllqf0i/weight_study.pdf?rlkey=17xykwkr8vw9rsxff7cg2dpjj&dl=0. Article: https://www.sonntagsblatt.de/corona-pandemie-lockdown-tu-ekfz-uebergewicht-hans-hauner.

Weight 2

"Germans Have Gotten Heavier Due to COVID-19"

Forty percent of Germans gained weight during the pandemic: around five to six kilograms. Those most affected are people aged 30 to 40 as well as socially disadvantaged families. According to experts, there are various reasons for the weight gain, including the "carbohydrate overload."

Many German citizens have apparently gained weight during the COVID-19 pandemic. The head of the Federal Center for Nutrition, which is part of the Ministry of Agriculture, Margareta Büning-Fesel, told the "Rheinische Post" (Thursday edition): "Germans have gotten heavier due to COVID-19."

Socially disadvantaged families have been particularly affected, said Büning-Fesel. Twenty-three percent of children from less-educated households have gained weight.

Among adults, higher weight gain was also observed: "Forty percent gained an average of five to six kilograms during the pandemic. Five COVID kilos is a lot." This primarily affected those aged 30 to 40, "especially those who were already overweight beforehand."

...

Original study: https://www.dropbox.com/scl/fi/rwyenyzslphws5gllqf0i/weight_study.pdf?rlkey=17xykwkr8vw9rsxff7cg2dpjj&dl=0. *Article:* https://www.welt.de/gesundheit/article236653313/Gewichtszunahme-Durch-Corona-sind-die-Deutschen-dicker-geworden.html.

Weight 3

"COVID-19 Fuels the Obesity Pandemic": 40 Percent Gained Weight

According to a new survey, a large portion of Germans gained a significant amount of weight during the pandemic. Those who were already overweight often gained even more. The main culprit is often a lack of physical activity.

It was somehow predictable: closed sports clubs and swimming pools, no commuting to work due to home office, fewer trips to other recreational activities — the measures to protect against the coronavirus also led to a loss of exercise for many people. And that has consequences: as a new survey commissioned by the Technical University of Munich now shows, about 40 percent of respondents gained weight.

Particularly concerning: 53 percent of those who already had weight problems before the pandemic reported gaining even more weight.

Respondents reported gaining an average of 5.6 kilograms since the beginning of the pandemic. Those with a body mass index over 30 even gained 7.2 kilograms. Regarding the causes, 33 percent said they had more time to eat. Twenty-eight percent said they ate more frequently out of boredom. More than half of the respondents said they had been less physically active since the beginning of the pandemic. Fifty-four percent of this subgroup said they moved less in their daily lives, and 53 percent attributed it to the closure of gyms and swimming pools.

...

Original study: https://www.dropbox.com/scl/fi/rwyenyzslphws5gllqf0i/weight_study.pdf?rlkey=17xykwkr8vw9rsxff7cg2dpjj&dl=0. *Article:* https://www.mdr.de/wissen/corona-pandemie-vierzig-prozent-der-deutschen-haben-zugenommen-100.html.

C.7 Prompts used for the classification by LLMs

C.7.1 Follow-Up Classification

Below you can find the prompts used for the classification of the accuracy of the follow-up article pool using ChatGPT's deep research function. The exercise we prompt the software with closely mirrors the training our research assistants received through the codebook. As, by design, the ChatGPT always asks a follow-up question after the first prompt the two prompts below were used in each iteration. The prompts were given in German, because the articles that had to be classified and the codebook are in German as well. A translation of to English is attached below.

Prompt 1 Betrachte die hier angehängte Excel-Tabelle. Deine Aufgabe ist es die Spalten "mistake" und "mistake_type" zu füllen. "mistake" ist ein binärer Wert, er wird 1 wenn in dem Artikel mindestens ein Fehler bei der Darstellung von Studienergebnissen passiert und 0, wenn das nicht der Fall ist. In "mistake_type" wird die Art des Fehlers spezifiziert. Dies kann eine der folgenden Kategorien sein:

- "causationcorrelation": Korrelation vs. Kausalität, dieser Fehler besteht wenn der Artikel einen Ursache-Wirkungs-Zusammenhang suggeriert, die Autoren in der Studie aber lediglich eine Korrelation beweisen. Beispiele für die Suggestion von Kausalität: "X führt zu Y", Wenn "X", dann "Y", Menschen mit "X" sind "Y"

- "context": Fehlender Kontext . Das ist ein Fehler, wenn der Kontext der Studie (z.B. Zeit/Ort) im Artikel nicht genannt wird, in der Studie aber als relevant hervorgehoben wird. Beispiele: Studie fand zu bestimmtem Zeitpunkt statt, der relevant ist (z.B. während der Corona-Pandemie), Studie fand nur in bestimmten Regionen statt, die für die Ergebnisse maßgeblich sind

- "sample": Stichprobe nicht repräsentativ. Das ist ein Fehler, wenn der Artikel nicht erwähnt, für wen oder was die Stichprobe repräsentativ ist oder wenn der Artikel nicht erwähnt, wer oder was genau die Stichprobe ist / nach welchen Merkmalen ausgewählt wurde. Auch wenn der Artikel nicht erwähnt, dass die Stichprobe möglicherweise zu klein ist, um allgemeine Schlüsse zu ziehen, besteht dieser Fehler. Beispiele: Aus 20 Beobachtungen wird eine allgemeine Aussage abgeleitet., Es wurden nur Menschen einer bestimmten Altersgruppe untersucht, das wird im Artikel aber nicht erwähnt.

- "statistics": Statistik / Kennzahl mathematisch nicht verstanden . Das ist ein Fehler, wenn z.B. der Artikel absolute und relative Wahrscheinlichkeiten

/ Häufigkeiten verwechselt, Prozent und Prozentpunkte verwechselt werden, bedingte Metriken (Häufigkeiten, Wahrscheinlichkeiten etc.) nicht korrekt wiedergegeben werden oder wenn Mittelwerte verallgemeinert werden. Beispiele: Bedingter Mittelwert wird als Mittelwert bezogen auf Population interpretiert., Statt einer Veränderung um X Prozentpunkte wird von einer Veränderung um X% gesprochen., Alle Tage im Juni sind sonniger als Tage im Dezember (richtig wäre: Im Schnitt sind Tage im Juni sonniger als im Dezember)

- “other": Eine andere Art von Fehler die nicht in die obigen Kategorien passen.

Du findest die jeweiligen Artikel auf die sich das bezieht mit den Links in der Spalte “article". Du musst dann den Artikel lesen und die Stelle finden, in der sich auf eine Studie bezogen wird. Dann musst du die Originalstudie suchen und die Ergebnisse mit der Darstellung im Artikel vergleichen. Wenn ein Fehler besteht codierst du das wie oben beschrieben.

Das soll nacheinander für alle Zeilen geschehen. Wenn es in einem Artikel Bezüge auf mehr als eine Studie gibt nimm die Studie, auf die sich am prominentesten bezogen wird. Das kann zB die Studie sein die am häufigsten genannt wird oder die zuerst angeführt wird. Wenn es mehr als einen Fehler gibt nimm den prominentesten Fehler.

Als Ergebnisse möchte ich einmal alle Zeilen hier angezeigt bekommen als Tabelle. Stelle sicher, dass du wirklich alle Ergebnisse präsentierst. Außerdem sollten in der Ergebnisdatei alle Original-Spalten, auch “ID" enthalten sein.

Prompt 2 Beziehe gerne alles ein.

Translation to English: Prompt 1 Consider the attached Excel sheet. Your task is to fill the columns “mistake” and “mistake_type”. “mistake” is a binary value: set it to 1 if the article contains at least one error in the presentation of study results, and to 0 if this is not the case. In “mistake_type” specify the kind of error. This can be one of the following categories:

- “causationcorrelation”: Correlation vs. causation. This error occurs when the article suggests a cause–effect relationship, but the authors of the study merely demonstrate a correlation. Examples of suggesting causality: “X leads to Y”, “If X, then Y”, “People with X are Y”.

- “context”: Missing context. This is an error when the context of the study (e.g. time/place) is not mentioned in the article, but is highlighted as relevant in the study. Examples: The study was conducted at a specific point in time that is relevant (e.g. during the COVID-19 pandemic); the study was carried out only in certain regions that are decisive for the results.

- “sample”: Non-representative sample. This is an error if the article does not mention for whom or what the sample is representative, or if the article does not mention who or what exactly the sample is / according to which criteria it was selected. This error also exists if the article does not mention that the sample may be too small to draw general conclusions. Examples: A general statement is derived from 20 observations; only people of a certain age group were examined but this is not mentioned in the article.

- “statistics”: Statistic / metric mathematically misunderstood. This is an error when, for example, the article confuses absolute and relative probabilities/frequencies, percent and percentage points, misrepresents conditional metrics (frequencies, probabilities, etc.), or generalises means. Examples: A conditional mean is interpreted as a mean with respect to the population; instead of a change of X percentage points the article speaks of a change of X%; “All days in June are sunnier than days in December” (correct would be: “On average, days in June are sunnier than in December”).

- “other”: Another type of error that does not fit into the categories above.

You can find the respective articles in question via the links in the column “article”. You must then read the article and locate the passage that refers to a study. Next, search for the original study and compare its results with the way they are presented in the article. If an error exists, code it as described above.

Carry this out sequentially for all rows. If an article refers to more than one study, choose the study that is referenced most prominently, e.g. the study that is mentioned most often or first. If there is more than one error, choose the most prominent error.

As output I would like to see a table with all rows displayed here. Make sure you really present all results. In addition, the result file should contain all original columns, including “ID”.

Translation to English: Prompt 2 Feel free to include everything.

C.7.2 Prompt for Headline Classification

Below you can find the prompts used for the classification of the headline data using the API version of Openais 4o model.

```
You are an expert news-linguistics scorer.
You will receive <headline>  </headline>.
Return **only** strict JSON with the seven keys
    sentiment, emotionality, specificity, newsworthy,
    clickbait, density, extreme
and no other text.

 Definitions
sentiment          integer (-1 negative, 0 neutral, 1 positive)

emotionality       float 01
   Based on Gennaro & Ash: higher values reflect a greater share of affective
    language
  versus cognitive language.

specificity        float 01
   Measures how narrowly the headlines subject is defined.
    0 = extremely broad or vague; 1 = extremely narrow, concrete and detailed.

newsworthy         float 01
   Uses Armona-Gentzkow-Kamenica-Shapiro concept of newsworthiness: joint
    measure of
  surprise and societal consequence for a general audience.
    0 = trivial or expected; 1 = highly surprising and consequential.

clickbait          float 01
   Captures curiosity-gap or sensational framing.
    0 = purely factual, literal; 1 = heavily curiosity-driven or sensational.

density            float 01
   Information content per word.
    0 = sparse, filler language; 1 = packed with concrete facts, numbers, named
    entities.

extreme            0 / 1 dummy
   1 only if the wording uses extremist, inflammatory, all-caps hyperbole, slurs,
  or similar extreme language.

 Output contract
 Reply **only** with the JSON object (no markdown, code fences, or extra prose).
 All numbers must be bare JSON numbers (not strings).
 Values must lie in their prescribed ranges.
```

```
If uncertain, choose the most reasonable value; do not output null.
```

C.8 Translation of the survey flow to English

As our participants are German journalists, the survey experiment was implemented in German language. Below we provide a translation of the survey flow into English. A dashed line indicates the next page of the survey. Explanatory comments (which were not displayed to participants) are indicated by blue, italic text.

Welcome Page

Welcome!

In this survey, we would like to ask for your assessment of the quality of different journalistic articles. Answering all the questions will take approximately 20 minutes. For your participation, we can offer you 10 euros as compensation.

All information provided here will be stored anonymously and used solely for research purposes. You can withdraw from participating in the survey at any time without giving any reason. However, to use your responses for research, it is important that you complete all the questions.

Involved Parties

This is a study conducted in collaboration with the Chair of Media Economics at the University of Cologne and the Center for Economic Studies at LMU Munich and the ifo Center for Industrial Organization and New Technologies at the ifo Institute in Munich. The principal investigators are Dr. Anna Kerkhof, Lara Berger, and Nikola Noske.

Funding

The survey is financially supported by the Chair of Media Economics at the University of Cologne, the Joachim Herz Foundation, and the Bavarian Research Institute for Digital Transformation (bidt).

Handling of Payment Data

To compensate you, we will ask for your PayPal address or bank account details at the end of the survey. This information will be stored separately from your other data and deleted after the payment is made. You also have the option to donate your compensation to an organization of your choice or to waive payment.

Use of Personal Data

For additional surveys, we will also ask for the name under which you publish articles in the media. We intend to use this name once to automatically analyze articles you have written and then store these results separately from your name. Your name will be deleted immediately after this analysis. If you do not wish to provide your name, you can leave the field blank.

You will find detailed information about data protection on the next page.

If you have any questions during the survey, feel free to contact lara.berger@uni-koeln.de at any time.

Informed Consent

Survey Information and Data Protection

General Information

The aim of this survey is to understand how journalists evaluate different science journalism articles. Data collection is conducted via an online questionnaire. Researchers from the University of Cologne and the Ludwig Maximilian University of Munich are involved in the survey.

Responsible Parties and Contact Information

The survey is primarily conducted by the University of Cologne. All information provided will be anonymized. If you have questions about the project or data processing, if you wish to withdraw your consent, or if you want to exercise other rights under the General Data Protection Regulation (GDPR), please contact Lara Berger. Email: lara.berger@uni-koeln.de

Details on Data Processing and Data Protection

Participants could click on a button and read a very detailed data protection statement, which is excluded from this translation in the interest of brevity.

Consent Statement

I have received and read the information sheet for the research project, and I consent to participate in this research project and the associated processing of my personal data. I am aware that I may decline to participate and that non-participation in the research project will not result in any disadvantages for me. I have been expressly informed of my right to withdraw consent.

- I consent.
- I do not consent.

Participants were excluded from further participation if they selected “I do not consent.”.

Job Status

Are you currently working as a journalist and/or have you been engaged in journalistic activities in the past 12 months?

- Yes
- No

Participants were excluded from further participation if they selected “No”.

Treatment: Video

Before we ask you further questions, we would like to share some tips on reporting scientific studies with you through this video. Please watch it in full. The button to proceed will appear after three minutes below.

After this text the video was embedded. A English transcript of the treatment video is available in Appendix C.5. This screen was only displayed to participants in the treatment group.

Instructions

In the following, we will show you articles that report on research findings.

Your task is to assess how accurately each text conveys the results of the studies. If you wish, you can access the original studies via a link.

In total, we will show you articles about two different studies.

Articles

Participants were randomly exposed to 2 out of a total of 14 news reports on scientific studies and a link to the original study. The transcripts of those articles are available in Appendix C.6. After each report participants were asked the following questions (same screen).

Were the statistics or results of the original study incorrectly or misleadingly represented in the article?

- Yes
- No

If, in your opinion, the statistics or results of the original study were incorrectly or misleadingly represented in the article: How many such errors are present in the text? (If there are no errors, please answer with "0").

— open text field that only accepted integers as reply —

If you think there are errors in the text: What type(s) of error(s) are they?

- Statistic/figure was not understood mathematically
- Missing context
- Confusion between correlation and causation
- Sample is not representative of the group being generalized
- Other
- There are no errors in the text.

Participants could select multiple replies.

If you had to write an article about the study, what headline would you choose?

— open text field —

After each article participants were exposed to one of the following attention checks. If participants failed an attention check they were excluded from further participation.

Attention Check 1

In surveys like this, it unfortunately happens quite often that participants do not read the questions carefully. To show us that you are paying attention, please answer the following question with "blue."

Based on what you have just read, what is your **favorite color**?

- Red
- Blue
- Green
- Yellow
- Orange
- Purple

Attention Check 2

In surveys like this, it unfortunately happens quite often that participants do not read the questions carefully. To show us that you are paying attention, please answer the following question with "green."

Based on what you have just read, what is your **favorite color**?

- Red
- Blue
- Green
- Yellow
- Orange
- Purple

Attention Check 3

In surveys like this, it unfortunately happens quite often that participants do not read the questions carefully. To show us that you are paying attention, please answer the following question with "dog."

Based on what you have just read, what is your **favorite animal**?

- Horse
- Cat
- Dog
- Mouse
- Bird
- Other animal

Attention Check 4

In surveys like this, it unfortunately happens quite often that participants do not read the questions carefully. To show us that you are paying attention, please answer the following question with "bird."

Based on what you have just read, what is your **favorite animal**?

- Horse
- Cat
- Dog
- Mouse
- Bird
- Other animal

Attention Check 5

In surveys like this, it unfortunately happens quite often that participants do not read the questions carefully. To show us that you are paying attention, please answer the following question with "mouse."

Based on what you have just read, what is your **favorite animal**?

- Horse
- Cat

- Dog
- Mouse
- Bird
- Other animal

Demographics

Thank you! You have almost reached the end of the survey.

Finally, we have a few questions about yourself.

How old are you?

— open text field that only accepted integers as reply —

Which gender do you identify with?

- male
- female
- diverse
- prefer not to say

How much money do you have available per month? (net)

- Less than 1500 euros

- 1500 euros - 2500 euros
- 2500 euros - 3500 euros
- 3500 euros - 4500 euros
- More than 4500 euros
- Prefer not to say

If the federal election were next Sunday, which party would you vote for?

- CDU/CSU
- SPD
- The Greens
- FDP
- AfD
- The Left
- Other
- None / I would not vote
- Prefer not to say

To complete our data, we would like to anonymously collect information on the type of content you publish and the type of media you work in. For this, we would need to record your name and the media organizations you mainly work for. Both your name and your employer(s) will be treated confidentially and will be deleted immediately once the data is collected.

If you do not wish to share this information, simply leave the fields blank.

Under what name do you usually publish articles in the media?

— open text field —

Which media outlet(s) do you usually work for?

— open text field —

Finally, we would like to ask for your feedback on this study. This will help us improve our research.

For example, were there any sections of this study that you did not understand or where you would have liked more detailed explanations?

— open text field —

Video Ex-Post

This page was only displayed to participants in the control-group. Clicking on [here] lead to our treatment video on YouTube.

Finally, we would like to provide you with a link to a video with some tips for reporting on scientific studies. If you wish, you can access the video [here].

Bibliography

Acemoglu, D., A. Ozdaglar, and J. Siderius (2025). "AI and Social Media: A Political Economy Perspective". In: MIT-Sloan Working Paper.

Alesina, A., A. Miano, and S. Stantcheva (2023). "Immigration and Redistribution". In: *The Review of Economic Studies* 90.1, pp. 1–39.

Ali, S., M. H. Saeed, E. Aldreabi, J. Blackburn, E. De Cristofaro, S. Zannettou, and G. Stringhini (2021). "Understanding the effect of deplatforming on social networks". In: *Proceedings of the 13th ACM Web Science Conference 2021*, pp. 187–195.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). "The Welfare Effects of Social Media". In: *American Economic Review* 110.3, pp. 629–76.

Allcott, H. and M. Gentzkow (2017). "Social Media and Fake News in the 2016 Election". In: *Journal of Economic Perspectives* 31.2, pp. 211–236.

Allen, J., D. J. Watts, and D. G. Rand (2024). "Quantifying the impact of misinformation and vaccine-skeptical content on Facebook". In: *Science* 384.6699, pp. 1–8.

Allen, J., A. A. Arechar, G. Pennycook, and D. G. Rand (2021). "Scaling up fact-checking using the wisdom of crowds". In: *Science Advances* 7.36, pp. 1–10.

Angelucci, C. and A. Prat (2024). "Is Journalistic Truth Dead? Measuring How Informed Voters Are about Political News". In: *American Economic Review* 114.4, pp. 887–925.

Arango-Kure, M., M. Garz, and A. Rott (2014). "Bad News Sells: The Demand for News Magazines and the Tone of Their Covers". In: *Journal of Media Economics* 27.4, pp. 199–214.

Armona, L., M. Gentzkow, E. Kamenica, and J. M. Shapiro (2024). "What is Newsworthy? Theory and Evidence". In: *Working Paper*. Forthcoming.

Bailey, R. L., J. J. Gahche, P. E. Miller, P. R. Thomas, and J. T. Dwyer (2013). "Why US adults use dietary supplements". In: *JAMA Internal Medicine* 173.5, pp. 355–361.

Bak-Coleman, J. B., I. Kennedy, M. Wack, A. Beers, J. S. Schafer, E. S. Spiro, K. Starbird, and J. D. West (2022). "Combining interventions to reduce the spread of viral misinformation". In: *Nature Human Behaviour* 6.10, pp. 1372–1380.

Balbuzanov, I., J. Gars, M. Stalinski, and E. Tjernström (2025). “Incentivizing Engagement: Experimental Evidence on Journalist Performance Pay”. In: Working paper.

Barrera, O., S. Guriev, E. Henry, and E. Zhuravskaya (2020). “Facts, alternative facts, and fact checking in times of post-truth politics”. In: *Journal of Public Economics* 182, p. 104123.

Berger, L. (2025). “How digital media markets amplify news sentiment”. In: *Working Paper*.

Berger, L. M., A. Kerkhof, F. Mindl, and J. Münster (2025). “Debunking “fake news” on social media: Immediate and short-term effects of fact-checking and media literacy interventions”. In: *Journal of Public Economics* 245, p. 105345.

Berger, L. M., A. Kerkhof, and N. Noske (2026). “Improving Science Literacy in the Newsroom: Experimental Evidence”. In: *PNAS Nexus* 5.4, pgag099.

Blair, G. and K. Imai (2012). “Statistical analysis of list experiments”. In: *Political Analysis* 20.1, pp. 47–77.

Bleich, E. and A. M. van der Veen (2021). “Media portrayals of Muslims: a comparative sentiment analysis of American newspapers, 1996–2015”. In: *Politics, Groups, and Identities* 9.1, pp. 20–39.

Bode, L. and E. K. Vraga (2015). “In related news, that was wrong: The correction of misinformation through related stories functionality in social media”. In: *Journal of Communication* 65.4, pp. 619–638.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2017). “Greater Internet use is not associated with faster growth in political polarization among US demographic groups”. In: *Proceedings of the National Academy of Sciences* 114.40, pp. 10612–10617.

— (2020). “Cross-Country Trends in Affective Polarization”. In: Working Paper 26669.

Brashier, N. M., G. Pennycook, A. J. Berinsky, and D. G. Rand (2021). “Timing matters when correcting fake news”. In: *Proceedings of the National Academy of Sciences* 118.5.

Broniatowski, D. A., J. R. Simons, J. Gu, A. M. Jamison, and L. C. Abroms (2023). “The efficacy of Facebook’s vaccine misinformation policies and architecture during the COVID-19 pandemic”. In: *Science Advances* 9.37.

Bursztyn, L., B. R. Handel, R. Jiménez-Durán, and C. Roth (2024). “When Product Markets Become Collective Traps: The Case of Social Media”. In: *American Economic Review*. Forthcoming.

Bursztyn, L., A. Rao, C. Roth, and D. Yanagizawa-Drott (2023). “Opinions as Facts”. In: *Review of Economic Studies* 90.4, pp. 1832–1864.

Cacciatore, M. A. (2021). “Misinformation and public opinion of science and health: Approaches, findings, and future directions”. In: *Proceedings of the National Academy of Sciences* 118.15, e1912437117.

Carey, J. M., A. M. Guess, P. J. Loewen, E. Merkley, B. Nyhan, J. B. Phillips, and J. Reifler (2022). “The ephemeral effects of fact-checks on COVID-19 misperceptions in the United States, Great Britain and Canada”. In: *Nature Human Behaviour* 6.2, pp. 236–243.

Carroll, C. D. (2003). “Macroeconomic Expectations of Households and Professional Forecasters”. In: *The Quarterly Journal of Economics* 118.1, pp. 269–298.

Chahrour, R., K. Nimark, and S. Pitschner (2021). “Sectoral Media Focus and Aggregate Fluctuations”. In: *American Economic Review* 111.12, pp. 3872–3922.

Cheung, A. C. and R. E. Slavin (2016). “How methodological features affect effect sizes in education”. In: *Educational Researcher* 45.5, pp. 283–292.

Chiou, W.-B., C.-C. Yang, and C.-S. Wan (2011). “Ironic effects of dietary supplementation: illusory invulnerability created by taking dietary supplements licenses health-risk behaviors”. In: *Psychological Science* 22.8, pp. 1081–1086.

Chopra, F., I. Haaland, F. Roeben, C. Roth, and V. Sticher (2025). “News Customization with AI”. In: CESifo Working Paper No. 12121.

Clarke, D., J. P. Romano, and M. Wolf (2020). “The Romano–Wolf multiple-hypothesis correction in Stata”. In: *The Stata Journal* 20.4, pp. 812–843.

Dahlstrom, M. F. (2021). “The narrative truth about scientific misinformation”. In: *Proceedings of the National Academy of Sciences* 118.15, e1914085117.

Dai, Y., W. Yu, and F. Shen (2021). “The effects of message order and debiasing information in misinformation correction”. In: *International Journal of Communication* 15, pp. 1039–1059.

De Fiore, F., A. Maurin, A. Mijakovic, and D. Sandri (2024). “Monetary policy in the news: communication pass-through and inflation expectations”. In: 1231. BIS Working Paper No. 1231.

De Quidt, J., J. Haushofer, and C. Roth (2018). “Measuring and bounding experimenter demand”. In: *American Economic Review* 108.11, pp. 3266–3302.

DellaVigna, S. and M. Gentzkow (2010). “Persuasion: Empirical Evidence”. In: *Annual Review of Economics* 2.1, pp. 643–669.

DellaVigna, S. and E. Kaplan (2007). “The Fox News effect: Media bias and voting”. In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

Dempster, E., R. Sutherland, and S. Keogh (2022). “Scientific Research in News Media: A Case Study of Misrepresentation, Sensationalism and Harmful Recommendations”. In: *Journal of Science Communication* 21.1, A06.

Dertwinkel-Kalt, M., J. Münster, and D. Zegners (2022). "If it Bleeds, it Leads: Negativity in Online News". In: *Working Paper*.

Deslauriers, L., E. Schelew, and C. Wieman (2011). "Improved learning in a large-enrollment physics class". In: *Science* 332.6031, pp. 862–864.

Drexler, A., G. Fischer, and A. Schoar (2014). "Keeping it simple: Financial literacy and rules of thumb". In: *American Economic Journal: Applied Economics* 6.2, pp. 1–31.

Drolsbach, C., K. Solovev, and N. Pröllochs (2024). "Community notes increase trust in fact-checking on social media". In: *PNAS Nexus* 3.7, pgae217.

Durante, R. and E. Zhuravskaya (2018). "Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict." In: *Journal of Political Economy* 126.3, pp. 1085–133.

Ecker, U. K. H., S. Lewandowsky, E. P. Chang, and R. Pillai (2014). "The effects of subtle misinformation in news headlines". In: *Journal of Experimental Psychology: Applied* 20.4, pp. 323–335.

Ecker, U. K., S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga, and M. A. Amazeen (2022). "The psychological drivers of misinformation belief and its resistance to correction". In: *Nature Reviews Psychology* 1.1, pp. 13–29.

Eisensee, T. and D. Strömberg (2007). "News Droughts, News Floods, and U.S. Disaster Relief". In: *The Quarterly Journal of Economics* 122.2, pp. 693–728.

Enikolopov, R., M. Petrova, and E. Zhuravskaya (2011). "Media and Political Persuasion: Evidence from Russia". In: *American Economic Review* 101.7, pp. 3253–3285.

Erbaugh, J. T., C. H. Chang, Y. J. Masuda, and J. Ribot (2024). "Communication and Deliberation for Environmental Governance". In: *Annual Review of Environment and Resources* 49, pp. 367–393.

Ershov, D. and J. S. Morales (2024). "Sharing News Left and Right: Frictions and Misinformation on Twitter". In: *The Economic Journal* 134.662, pp. 2391–2417.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). "Global Evidence on Economic Preferences*". In: *The Quarterly Journal of Economics* 133.4, pp. 1645–1692.

Farnsworth, S. and S. R. Lichter (2011). "The nightly news nightmare". In: *Lanham MD: Rowman Littlefield.*

Farrell, H. (2012). "The Consequences of the Internet for Politics". In: *Annual Review of Political Science* 15.1, pp. 35–52.

Fong, J., T. Guo, and A. Rao (2024). "Debunking Misinformation About Consumer Products: Effects on Beliefs and Purchase Behavior". In: *Journal of Marketing Research* 61.4, pp. 659–681.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth (2014). "Active learning increases student performance in science, engineering, and mathematics". In: *Proceedings of the National Academy of Sciences* 111.23, pp. 8410–8415.

Fryer Jr, R. G. (2017). "The production of human capital in developed countries: Evidence from 196 randomized field experiments". In: *Handbook of economic field experiments*. Vol. 2. Elsevier, pp. 95–322.

Gal, I. (2002). "Adults' Statistical Literacy: Meanings, Components, Responsibilities". In: *International Statistical Review* 70.1, pp. 1–25.

Gambetti, L., N. Maffei-Faccioli, and S. Zoi (2023). "Bad News, Good News: Coverage and Response Asymmetries". In: 2023-001.

Garg, P. and T. Fetzer (2025). "Political expression of academics on Twitter". In: *Nature Human Behaviour*.

Garz, M. (2014). "Good news and bad news: evidence of media bias in unemployment reports". In: *Public Choice* 161, pp. 499–515.

Gennaro, G. and E. Ash (2022). "Emotion and Reason in Political Language". In: *The Economic Journal* 132.643, pp. 1037–1059.

Gensing, P. (2021). "Wie die AfD Angst vor Impfungen schürt". In: *Tagesschau.de*.

Gentzkow, M. and J. M. Shapiro (2010). "What Drives Media Slant? Evidence from U.S. Daily Newspapers". In: *Econometrica* 78.1, pp. 35–71.

— (2011). " Ideological Segregation Online and Offline *". In: *The Quarterly Journal of Economics* 126.4, pp. 1799–1839.

Goidel, R. K. and R. E. Langley (1995). "Media Coverage of the Economy and Aggregate Economic Evaluations: Uncovering Evidence of Indirect Media Effects". In: *Political Research Quarterly* 48.2, pp. 313–328.

Graves, L., B. Nyhan, and J. Reifler (2016). "Understanding innovations in journalistic practice: A field experiment examining motivations for fact-checking". In: *Journal of Communication* 66.1, pp. 102–138.

Greiner, B. (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE". In: *Journal of the Economic Science Association* 1.1, pp. 114–125.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). "Fake news on Twitter during the 2016 US presidential election". In: *Science* 363.6425, pp. 374–378.

Gu, J., A. Dor, K. Li, D. A. Broniatowski, M. Hatheway, L. Fritz, and L. C. Abroms (2022). "The impact of Facebook's vaccine misinformation policy on user endorsements of vaccine content: An interrupted time series analysis". In: *Vaccine* 40.14, pp. 2209–2214.

Guay, B., A. J. Berinsky, G. Pennycook, and D. Rand (2023). "How to think about whether misinformation interventions work". In: *Nature Human Behaviour* 7.8, pp. 1231–1233.

Guess, A., J. Nagler, and J. Tucker (2019). "Less than you think: Prevalence and predictors of fake news dissemination on Facebook". In: *Science Advances* 5.1, eaau4586.

Guess, A., B. Nyhan, and J. Reifler (2018). "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign". In: *European Research Council* 9.3, p. 4.

Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar (2020). "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India". In: *Proceedings of the National Academy of Sciences* 117.27, pp. 15536–15545.

Gundersen, T., D. Alinejad, T. Y. Branch, B. Duffy, K. Hewlett, C. Holst, S. Owens, F. Panizza, S. M. Tellman, J. van Dijck, and M. Baghramian (2022). "A new dark age? Truth, trust, and environmental science". In: *Annual Review of Environment and Resources* 47.1, pp. 5–29.

Guriev, S., E. Henry, T. Marquis, and E. Zhuravskaya (2023). "Curtailing False News, Amplifying Truth". In: *Working Paper*.

Hanitzsch, T., N. Steindl, and C. Lauerer (2016). "Country Report: Journalists in Germany". In: *Worlds of Journalism Study*.

Hartmann, J., M. Heitmann, C. Siebert, and C. Schamp (2022). "More than a Feeling: Accuracy and Application of Sentiment Analysis". In: *International Journal of Research in Marketing*.

Henkel, L. and C. Zimpelmann (2022). "Proud to Not Own Stocks: How Identity Shapes Financial Decisions". In: *ECONtribute Discussion Paper No. 206*.

Henry, E., E. Zhuravskaya, and S. Guriev (2022). "Checking and Sharing Alt-Facts". In: *American Economic Journal: Economic Policy* 14.3, pp. 55–86.

Hill, C. J., H. S. Bloom, A. R. Black, and M. W. Lipsey (2008). "Empirical benchmarks for interpreting effect sizes in research". In: *Child development perspectives* 2.3, pp. 172–177.

Hofstetter, C. R. and D. M. Dozier (1986). "Useful News, Sensational News: Quality, Sensationalism and Local TV News". In: *Journalism Mass Communication Quarterly* 63.4, pp. 815–53.

Hopkins, D. J. (2009). “No more Wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates”. In: *The Journal of Politics* 71.3, pp. 769–781.

Hutto, C. J. and E. E. Gilbert (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.*

Jerit, J. and Y. Zhao (2020). “Political misinformation”. In: *Annual Review of Political Science* 23, pp. 77–94.

Kaiser, T., A. Lusardi, L. Menkhoff, and C. Urban (2022). “Financial education affects financial knowledge and downstream behaviors”. In: *Journal of Financial Economics* 145.2, pp. 255–272.

Kaiser, T. and L. Menkhoff (2022). “Active learning improves financial education: Experimental evidence from Uganda”. In: *Journal of Development Economics* 157.

Karlin, B., J. F. Zinger, and R. Ford (2015). “The effects of feedback on energy conservation: A meta-analysis.” In: *Psychological Bulletin* 141.6, pp. 1205–1227.

Kayser, M. A. and M. Peress (2021). “Does the Media Cover the Economy Accurately? An Analysis of Sixteen Developed Democracies”. In: *Quarterly Journal of Political Science* 16.1, pp. 1–33.

Kirkegaard, E. O. W., J. Pallesen, E. Elgaard, and N. Carl (2021). “The Left-liberal Skew of Western Media”. In: *Journal of Psychological Research* 3.

Klamm, C., I. Rehbein, and S. Ponzetto (2022). “FrameASt: A Framework for Second-level Agenda Setting in Parliamentary Debates through the Lense of Comparative Agenda Topics”. In: *ParlaCLARIN III at LREC2022*.

Kuklinski, J. H., P. J. Quirk, J. Jerit, D. Schwieder, and R. F. Rich (2000). “Misinformation and the currency of democratic citizenship”. In: *The Journal of Politics* 62.3, pp. 790–816.

Larsen, V. H., L. A. Thorsrud, and J. Zhulanova (2021). “News-driven inflation expectations and information rigidities”. In: *Journal of Monetary Economics* 117, pp. 507–520.

Lazer, D. M., M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, and D. Rothschild (2018). “The science of fake news”. In: *Science* 359.6380, pp. 1094–1096.

Leung, T. C. and K. S. Strumpf (2023). “All the Headlines that Are Fit to Change”. In: *SSRN*. Working Paper.

Lewandowsky, S. (2021). “Climate change disinformation and how to combat it”. In: *Annual Review of Public Health* 42.1, pp. 1–21.

Lewandowsky, S., U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook (2012). "Misinformation and its correction: Continued influence and successful debiasing". In: *Psychological science in the public interest* 13.3, pp. 106–131.

Lewandowsky, S. and S. Van Der Linden (2021). "Countering misinformation and fake news through inoculation and prebunking". In: *European Review of Social Psychology* 32.2, pp. 348–384.

Liu, X., L. Qi, L. Wang, and M. J. Metzger (2023). "Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness". In: *Communication Research*.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*.

Loewenstein, G. and Z. Wojtowicz (2023). "The Economics of Attention". In: *Available at SSRN 4368304*.

Loosen, W., A. von Garmissen, E. Bartelt, and T. van Olphen (2023). *Journalismus in Deutschland 2023. Aktuelle Befunde zu Situation und Wandel*. Tech. rep. Arbeitspapiere des Hans-Bredow-Instituts, Projektergebnisse Nr. 68. Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI).

Loughran, T. and B. McDonald (2011). "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

Luca, M. (2015). "User-generated content and social media". In: *Handbook of Media Economics*. Vol. 1. Elsevier, pp. 563–592.

Maaß, S., J. Wortelker, and A. Rott (2024). "Evaluating the regulation of social media: An empirical study of the German NetzDG and Facebook". In: *Telecommunications Policy* 48.5, p. 102719.

Maertens, R., J. Roozenbeek, M. Basol, and S. van der Linden (2021). "Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments." In: *Journal of Experimental Psychology: Applied* 27.1, p. 1.

Markowitz, D. M. (2024). "From complexity to clarity: How AI enhances perceptions of scientists and the public's understanding of science". In: *PNAS Nexus* 3.9, p. 387.

Martel, C. and D. G. Rand (2024). "Fact-checker warning labels are effective even for those who distrust fact-checkers". In: *Nature Human Behaviour* 8, pp. 1957–1967.

Mehmood, S., S. Naseer, and D. L. Chen (2021). "Training policymakers in econometrics". In: *Working Paper*.

Meyer, T., A. Kerkhof, C. Cennamo, and T. Kretschmer (2024). "Competing for Attention on Digital Platforms: The Case of News Outlets". In: *Strategic Management Journal* 45.9, pp. 1731–1790.

Mitts, T., N. Pisharody, and J. Shapiro (2022). “Removal of anti-vaccine content impacts social media discourse”. In: *Proceedings of the 14th ACM Web Science Conference 2022*, pp. 319–326.

Mullainathan, S. and A. Shleifer (2005). “The Market for News”. In: *American Economic Review* 95.4, pp. 1031–1053.

Mummolo, J. and E. Peterson (2019). “Demand effects in survey experiments: An empirical assessment”. In: *American Political Science Review* 113.2, pp. 517–529.

Nickl, P., M. Moussaïd, and P. Lorenz-Spreen (2025). “The evolution of online news headlines”. In: *Humanities and Social Sciences Communications* 12.364.

Noar, S. M., C. N. Benac, and M. S. Harris (2007). “Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions.” In: *Psychological Bulletin* 133.4, p. 673.

Nyhan, B. (2021). “Why the backfire effect does not explain the durability of political misperceptions”. In: *Proceedings of the National Academy of Sciences* 118.15, e1912440117.

Nyhan, B., E. Porter, J. Reifler, and T. J. Wood (2020). “Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability”. In: *Political Behavior* 42.3, pp. 939–960.

Nyhan, B. and J. Reifler (2010). “When corrections fail: The persistence of political misperceptions”. In: *Political Behavior* 32.2, pp. 303–330.

— (2015). “Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information”. In: *Vaccine* 33.3, pp. 459–464.

Oxman, M., L. Larun, G. P. Gaxiola, D. Alsaid, A. Qasim, C. J. Rose, K. Bischoff, and A. D. Oxman (2022). “Quality of information in news media reports about the effects of health interventions: Systematic review and meta-analyses”. In: *F1000Research* 10, p. 433.

Pariser, E. (2012). *The Filter Bubble: How the New Personalized Web is Changing what We Read and how We Think*. Penguin Books.

Pennycook, G., Z. Epstein, M. Mosleh, A. A. Arechar, D. Eckles, and D. G. Rand (2021). “Shifting attention to accuracy can reduce misinformation online”. In: *Nature* 592.7855, pp. 590–595.

Pennycook, G., J. McPhetres, Y. Zhang, J. G. Lu, and D. G. Rand (2020). “Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention”. In: *Psychological Science* 31.7, pp. 770–780.

Pennycook, G. and D. G. Rand (2019). “Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning”. In: *Cognition* 188, pp. 39–50.

Pennycook, G. and D. G. Rand (2021). “The Psychology of Fake News”. In: *Trends in Cognitive Sciences* 25.5, pp. 388–402.

Radimer, K., B. Bindewald, J. Hughes, B. Ervin, C. Swanson, and M. F. Picciano (2004). “Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000”. In: *American Journal of Epidemiology* 160.4, pp. 339–349.

Rauscher, F. H., G. L. Shaw, and K. N. Ky (1993). “Music and spatial task performance”. In: *Nature* 365.6447, p. 611.

Remus, R., G. Heyer, and U. Quasthoff (2010). *SentiWS - a Publicly Available German-language Resource for Sentiment Analysis.*

Reporters Without Borders (2024). *World Press Freedom Index 2024.*

Robertson, C. E., N. Pröllochs, K. Schwarzenegger, P. Pärnamets, J. J. Van Bavel, and S. Feuerriegel (2023). “Negativity drives online news consumption”. In: *Nature Human Behaviour* 7.5, pp. 812–822.

Romano, J. P. and M. Wolf (2005). “Stepwise multiple testing as formalized data snooping”. In: *Econometrica* 73.4, pp. 1237–1282.

Rooney, B. L. and D. M. Murray (1996). “A meta-analysis of smoking prevention programs after adjustment for errors in the unit of analysis”. In: *Health education quarterly* 23.1, pp. 48–64.

Roozenbeek, J., S. van der Linden, B. Goldberg, S. Rathje, and S. Lewandowsky (2022). “Psychological inoculation improves resilience against misinformation on social media”. In: *Science Advances* 8.34, eabo6254.

Rozin, P. and E. B. Royzman (2001). “Negativity Bias, Negativity Dominance, and Contagion”. In: *Personality and Social Psychology Review* 5.4, pp. 296–320.

Ruiz-Primo, M. A., D. Briggs, H. Iverson, R. Talbot, and L. A. Shepard (2011). “Impact of undergraduate science course innovations on learning”. In: *Science* 331.6022, pp. 1269–1270.

Rumsey, D. J. (2002). “Statistical literacy as a goal for introductory statistics courses”. In: *Journal of Statistics Education* 10.3, pp. 1–17.

Ryu, J. S. (1982). “Public Affairs and Sensationalism in Local TV News Programs”. In: *Journalism Mass Communication Quarterly* 59.1, pp. 74–137.

Schield, M. (1999). “Statistical literacy: Thinking critically about statistics”. In: *Of Significance* 1.1, pp. 15–20.

Selvaraj, S., D. S. Borkar, and V. Prasad (2014). “Media coverage of medical journals: Do the best articles make the news?” In: *PLOS ONE* 9.1, e85355.

Serra-Garcia, M. (2025). *The Attention–Information Tradeoff.* CESifo Working Paper 11885. Available at SSRN. CESifo.

Shapiro, A. H., M. Sudhof, and D. J. Wilson (2022). “Measuring news sentiment”. In: *Journal of Econometrics* 228.2, pp. 221–243.

Sharma, S. (2017). “Definitions and models of statistical literacy: a literature review”. In: *Open Review of Educational Research* 4.1, pp. 118–133.

Simon, H. A. (1971). “Designing Organizations for an Information-Rich World”. In: *Computers, Communications, and the Public Interest*. Ed. by M. Greenberger. Baltimore, MD: The Johns Hopkins Press, pp. 38–72.

Snyder James M., J. and D. Strömberg (2010). “Press Coverage and Political Accountability”. In: *Journal of Political Economy* 118.2, pp. 355–408.

Soleimanian, M. (2022). *Do Firms Walk Their Talk in Corporate Social Responsibility Reports? – Evidence From Forward-Looking Statements*.

Soroka, S. (2016). “Gatekeeping and the Negativity Bias”. In: *Oxford Research Encyclopedia of Politics*.

Soroka, S., M. Daku, D. Hiaeshutter-Rice, L. Guggenheim, and J. Pasek (2018). “Negativity and Positivity Biases in Economic News Coverage: Traditional Versus Social Media”. In: *Communication Research* 45.7, pp. 1078–1098.

Soroka, S., P. Fournier, and L. Nir (2019). “Cross-national evidence of a negativity bias in psychophysiological reactions to news”. In: *Proceedings of the National Academy of Sciences* 116.38, pp. 18888–18892.

Soroka, S. and Y. Krupnikov (2021). *The Increasing Viability of Good News*. Elements in Politics and Communication. Cambridge University Press.

Stantcheva, S. (2021). “Understanding tax policy: How do people reason?” In: *The Quarterly Journal of Economics* 136.4, pp. 2309–2369.

Suleski, J. and M. Ibaraki (2010). “Scientists are talking, but mostly to each other: a quantitative analysis of research represented in mass media”. In: *Public Understanding of Science* 19.1, pp. 115–125.

Summ, A. and A.-M. Volpers (2016). “What’s science? Where’s science? Science journalism in German print media”. In: *Public Understanding of Science* 25.7, pp. 775–790.

Sumner, P., S. Vivian-Griffiths, J. Boivin, A. Williams, L. Bott, R. Adams, C. A. Venetis, L. Whelan, B. Hughes, and C. D. Chambers (2016). “Exaggerations and Caveats in Press Releases and Health-Related Science News”. In: *PLoS ONE* 11.12, e0168217.

Swire, B., A. J. Berinsky, S. Lewandowsky, and U. K. Ecker (2017). “Processing political misinformation: comprehending the Trump phenomenon”. In: *Royal Society Open Science* 4.3.

Swire-Thompson, B. and D. Lazer (2020). “Public health and online misinformation: challenges and recommendations”. In: *Annual Review of Public Health* 41.1, pp. 433–451.

Swire-Thompson, B., J. Cook, L. H. Butler, J. A. Sanderson, S. Lewandowsky, and U. K. Ecker (2021). “Correction format has a limited role when debunking misinformation”. In: *Cognitive Research: Principles and Implications* 6.83, pp. 1–15.

Tetlock, P. C. (2007). “Giving Content to Investor Sentiment: The Role of Media in the Stock Market”. In: *The Journal of Finance* 62.3, pp. 1139–1168.

Thompson, E. R. (2007). “Development and Validation of an Internationally Reliable Short-Form of the Positive and Negative Affect Schedule (PANAS)”. In: *Journal of Cross-Cultural Psychology* 38.2, pp. 227–242.

Trussler, M. and S. Soroka (2014). “Consumer Demand for Cynical and Negative News Frames”. In: *The International Journal of Press/Politics* 19.3, pp. 360–379.

Tversky, A. and D. Kahneman (1974). “Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.” In: *Science* 185.4157, pp. 1124–1131.

Verplanken, B. and S. Orbell (2022). “Attitudes, habits, and behavior change.” In: *Annual Review of Psychology* 73, pp. 327–352.

Vestergaard, G. L. and K. H. Nielsen (2016). “Science news in a closed and an open media market: A comparative content analysis of print and online science news in Denmark and the United Kingdom”. In: *European Journal of Communication* 31.6, pp. 661–677.

Vosoughi, S., D. Roy, and S. Aral (2018). “The spread of true and false news online”. In: *Science* 359.6380, pp. 1146–1151.

Vraga, E. K. and L. Bode (2017). “Using expert sources to correct health misinformation in social media”. In: *Science Communication* 39.5, pp. 621–645.

Watson, J., S. van der Linden, M. Watson, and D. Stillwell (2024). “Negative online news articles are shared more to social media”. In: *Scientific Reports* 14.1, p. 21592.

Willnat, L., D. H. Weaver, and G. C. Wilhoit (2022). “The American Journalist Under Attack: Key Findings 2022”. In: *Report S.I. Newhouse School of Public Communications*.

Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. 1st. London: Atlantic Books.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). “Political effects of the internet and social media”. In: *Annual Review of Economics* 12, pp. 415–438.