Debunking "fake news" on social media: short- and longer-term effects of fact checking and media literacy interventions*

Lara Marie Berger†
Anna Kerkhof‡
Felix Mindl§
Johannes Münster¶

This version: September 1, 2022

Abstract

We conduct a randomized survey experiment to compare the short- and longer-term effects of fact checking to a brief media literacy intervention. We show that the impact of fact checking is limited to the corrected fake news, whereas media literacy helps to distinguish between false and correct information more generally, both immediately and 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. Our results promote media literacy as an effective tool to fight fake news, that is cheap, scalable, and easy-to-implement.

JEL Codes: L51, L82, Z18

Keywords: Covid-19, Facebook, fact checking, fake news, media literacy, misinformation, social media, survey experiment, vaccine

*We thank Maja Adena, Kai Barron, Anna Bindler, Esther Blanco, Grazia Cecere, Felix Chopra, Oliver Falck, Marcel Garz, Julian Harke, Paul Hufe, Jonas Loebbing, Dominik Rehse, Heiner Schumacher, Andreas Steinmayr, Maiting Zhuang, and participants of various seminars and conferences for helpful comments and suggestions. Nikola Noske provided excellent research assistance. Anna Kerkhof acknowledges financial support from the Joachim Herz Stiftung. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2126/1-390838866. 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. All errors are our own.

†University of Cologne. E-mail: lara.berger[at]wiso.uni-koeln.de.

‡Corresponding author. University of Munich, ifo Institute for Economic Research Munich, bidt Graduate Center, and CESifo. E-mail: kerkhof[at]ifo.de

§University of Cologne and iwp Institute for Economic Policy. E-mail: mindl[at]wiso.uni-koeln.de

¶University of Cologne. E-mail: johannes.muenster[at]uni-koeln.de.

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., Lazer et al., 2018). Social media like Facebook and 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). Such numbers are alarming, especially during a global health crisis, where exposure to false or misleading information can be a matter of life and death.

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 Facebook and Twitter have started to flag suspicious content, too. Yet, it is unclear whether these remedies function as desired: empirical evidence on the effectiveness of fact checking is mixed (Vraga and Bode, 2017; Jerit and Zhao, 2020), knowledge on the impact of media literacy is scarce (Guess et al., 2020), and a direct comparison of these interventions does not exist at all.

We address this gap with a large-scale randomized survey experiment on the short- and longer-term effects of fact checking and media literacy interventions. In the experiment, we expose participants to false and correct statements on health-related topics – Corona vaccines and nutrition – that we retrieve from Facebook ("fakes" and "facts"). 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 training. Then, we compare the two treatment groups to participants who do not receive an intervention. To study longer-term effects, we re-invite the same participants about two weeks later to a second, analogous wave of the survey.

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 in the short- and in the longer-run. 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 enhance their skills.

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) in the short-run, but only the media literacy intervention reduces the credibility of fakes on nutrition (which are not corrected by fact checks), both in the short- and in the longer-run. Moreover, both interventions improve participants' factual knowledge

[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` (Aug 2022) for further information on efforts by the European Union.

in the short-, but only the media literacy intervention in the longer-run. Finally, while the media literacy intervention raises participants' willingness to get vaccinated (or boostered) against Covid-19 in the short- and in the longer-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 poor. In particular, both the fact checking and the media literacy intervention are *more* effective for supporters of the AfD ("Alternative for Germany", a far-right populist party known for spreading misinformation on Covid-19) for fakes on Corona vaccines, but not for fakes on nutrition, where participants' beliefs are much more alike. Computing persuasion rates à la DellaVigna and Kaplan (2007) shows that this result can just partly be explained by differences in the proportion of participants who are left to be convinced. However, we also provide evidence that AfD 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 reasonable mechanism for our results is that the media literacy intervention raises participants' attention and enables them to critically evaluate the Facebook postings, while fact checking fails to 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. Moreover, media literacy helps participants to better identify untrustworthy elements in fakes and trustworthy elements in facts, thus increasing their ability for truth discernment. Fact checking, in contrast, has no such effect.

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 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 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 both the fact checking and, in particular, the media literacy intervention can effectively repeal that impact.

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 debunk 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. We show that in an environment where only a small proportion of fake news can ever be fact checked, media literacy 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 (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. (2020, 2021) show that users share false claims partly because they do not think sufficiently about whether or not the content is accurate. 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 enhance users' skills – as less effective.

The remainder of the paper is organized as follows. Section 2 reviews the related literature on social media, user-generated content, and misinformation. Section 3 illustrates the experimental setup and implementation; moreover, we discuss our empirical analysis. Section 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 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 6 presents further results and robustness checks, Section 7 concludes.

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; Lazer et al., 2018; Guess et al., 2018, 2019; Grinberg et al., 2019; Vosoughi et al., 2018), and inquiries of potential remedies (reviewed by Lewandowsky et al., 2012; Jerit and Zhao, 2020). The latter literature focuses on corrective interventions like fact checking: While Bode and Vraga (2015), Vraga and Bode (2017), and Henry et al. (2020), among others, support its effectiveness, other papers find no or even "backfire" effects (e.g., Nyhan and Reifler, 2010, 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). Studies on alternative ways to combat fake news are rare. One notable exception are 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

the quality of their sharing decisions.

We contribute to this literature in several ways. First, we pursue a clean comparison of fact checking and media literacy interventions, which has not been done so far. In particular, our experimental setup allows us to study the short- and longer-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.

Most closely related to our study are Barrera et al. (2020) and 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.[2] 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 short- *and* longer-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 the external validity of our results extends to messages 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., Pennycook and Rand, 2019). Pennycook et al. (2020, 2021), 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 support 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., Lewandowsky et al., 2012; Jerit and Zhao, 2020), we find that our interventions are more effective for supporters of right-wing extremists, 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 that causes the ineffectiveness of fact checking.

Education interventions Second, our paper is related to the broad literature on education interventions.

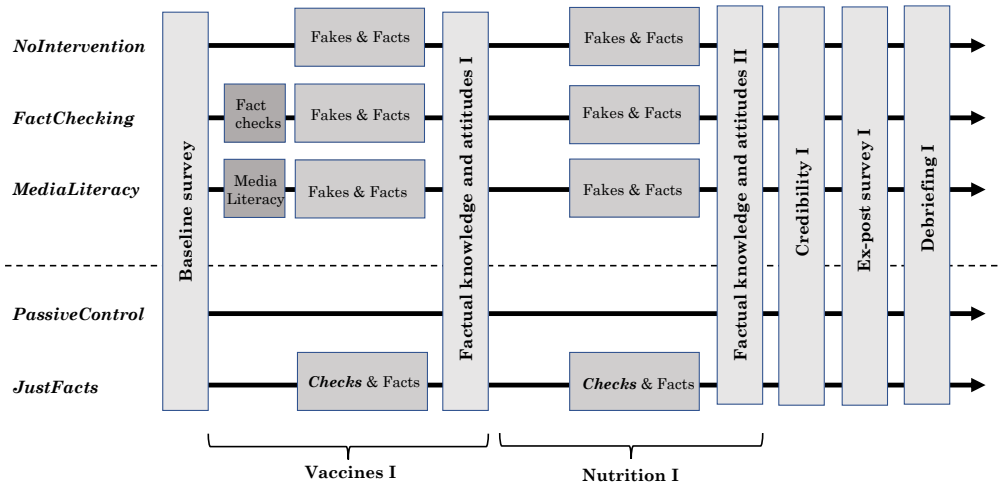
[2]Similar results are presented by Nyhan et al. (2020).

to be completed

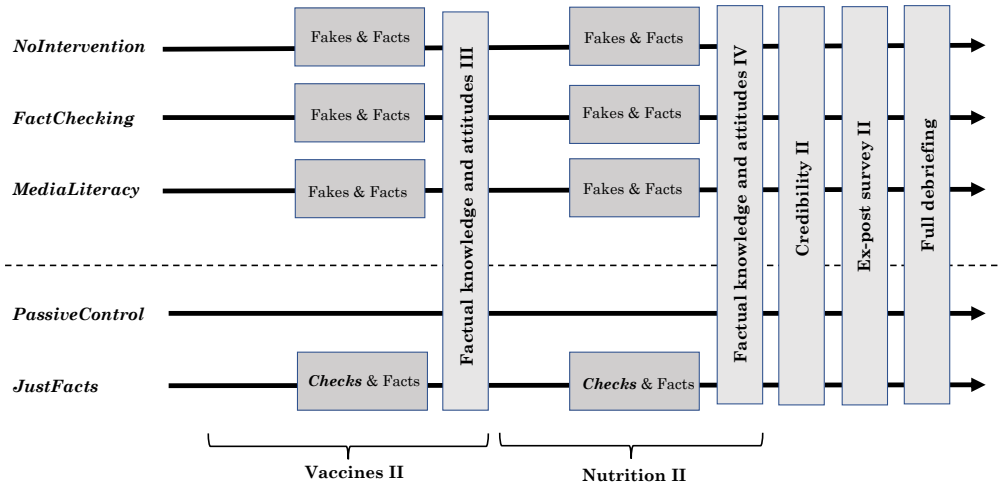
3. Experimental design

3.1. Survey flow

We start by randomizing the participants of our online survey experiment into one out of five groups of equal size: (i) NoIntervention, (ii) FactChecking, (iii) MediaLiteracy, (iv) JustFacts, and (v) PassiveControl. To study both the short- and longer-term effects of our interventions, we conduct two waves of the experiment, where we re-invite the same participants about one week after they completed Wave I and allocate them to the same treatment group as before. Figure 1 gives an overview of our survey flow, further details are discussed below.



(a) Overview Wave I



(b) Overview Wave II

Figure 1: Survey flow

3.1.1. Wave I

Baseline survey All participants start with a baseline survey on standard demographics like 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. To avoid priming effects on subsequent questions, we 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*.

Vaccines Next, 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.[3] The fakes and facts were manually collected from Facebook, i.e., we use screenshots of Facebook postings that actually exist. Fakes were only included if we could find appropriate fact checks that debunk the false information. Moreover, all fakes and facts must contain a concrete numerical value (e.g., "50 people died after vaccination in a Sana clinic") that we could later on ask for. See Appendix C.2 for all fakes and facts that we use.

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*, a major German fact checking initiative). 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. In addition, displaying the fact check prior to the respective fake makes the fact checking better comparable to the media literacy intervention. See Appendix C.2 for all fact checks that we use.

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.[4] These tips actually exist on the platform and comprise ten short pieces of advice, including "Be skeptical of headlines", "Look closely at the link", and "Investigate the source"; Appendix C.1 shows the full list. Participants are aware 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.

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.[5] We can thereby infer our

[3] 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.

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

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

participants' average prior beliefs and attitudes from responses by the PassiveControl, and the impact of stand-alone fact checks from the JustFacts group.

After exposure to the Facebook postings, we ask all participants four factual questions that are tailored to the fakes and facts just shown (e.g., we ask how many people died after vaccination in a Sana clinic).[6] 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.[7]

Next, we inquire all participants' willingness to get vaccinated against Covid-19. We start by partitioning participants into those who are already fully vaccinated by the time of the experiment, and those who are not.[8] 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 6.5.3 for further discussion).

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, the FactChecking, and the MediaLiteracy group). 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.[9] 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.[10]

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. 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*.

[6] The correct answer is zero.

[7] 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.

[8] The experiment was conducted in September / October 2021, when all German adults had had the opportunity to get fully vaccinated (two injections), and when policy makers and health experts were discussing whether and when a third injection would make sense.

[9] Though not as topical as Corona vaccines, the consumption of (needless) dietary supplements is an important concern. Recent surveys indicates 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).

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

Credibility Next, we inquire the perceived credibility of all fakes, facts, and fact checks. To this end, we display them all again and let participants rate their 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 serves to enhance our understanding of potential mechanisms as well as to collect information that we did not inquire earlier to avoid priming effects.

We proceed in three steps. First, we display all fakes, facts, and fact checks again and let participants indicate which elements of the postings they perceive as especially trustworthy or untrustworthy. We pre-define five types of elements to obtain comparable responses: (i) format and spelling, (ii) content as such, (iii) pictures, (iv) source and URL, and (v) verified account.[11] Participants can mark none, one, or several elements per posting and must complete a mandatory tutorial to get familiar with the technique before they can proceed (see Appendix C.3 for an example). As with the credibility questions, participants are only asked about postings that they saw during the experiment.

Second, we ask further questions on participants' political party preferences, social media usage, if they got vaccinated 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.

Third, we conduct two list experiments à la Blair and Imai (2012) to rebut that our main results on attitudes are driven by a "Bradley effect" (Hopkins, 2009), whereby participants conceal socially undesirable opinions and attitudes (see Section 6.5.3 for further details).

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.

3.1.2. Wave II

To study the effectiveness of our fact checking and media literacy interventions in the longer-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.

We conclude the survey with a full debriefing of all participants. To this end, we display the correct answers to all factual questions and provide links to trustworthy websites on Corona vaccines and nutrition, where the participants can get further information on these topics if they

[11] Note that some of the postings do not exhibit every element; e.g., some of them have no *verified account* label.

wish.[12]

3.2. Implementation

The experiment was programmed with the survey software *Qualtrics* and conducted in cooperation with *respondi*, a major commercial panel provider.[13] We used e-mails to invite around 3,000 participants to Wave I of the survey, i.e., around 600 participants per group. 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.[14] 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.

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 eights days, though, the median interval is equal to 15, and the mean interval equal to 17.6 days (see Section 6.5.2 for further discussion). The response rate is equal to 83% – which is roughly equal to *respondi*'s average – and we find no evidence for differential attrition.[15] At the time the experiment took place, all German adults had had the 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. Balance check

Table A.1 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.[16] 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 4.1.3) our results are unaffected.

[12] 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.

[13] Cooperating with professional panel providers such as *respondi* has become standard in economic research; see, e.g., Stantcheva (2021) and Alesina et al. (2022) for examples and `https://www.respondi.com/` for further details on *respondi*.

[14] 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.

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

[16] The pre-registered 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.

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 A.2 provides summary statistics of all dependent variables that we use in the analysis.

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 *Credible* or *Very credible*.[17] This aggregation level allows us to 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.[18] Then, 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 *Likely* or *Very likely* 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 *Unlikely* or *Very unlikely* to consume dietary supplements in the near future.[19]

Likes and dislikes 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 they marked as trustworthy ("like" for brevity) or untrustworthy ("dislike") in each posting. Then, we compute the absolute number of likes and dislikes 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 standardized distance

[17] Our results are robust to alternative cutoffs.

[18] Note that we pre-registered our intention to winsorize participants' 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.

[19] Our results are robust to alternative cutoffs.

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 *Undecided* about his or her prior knowledge.

3.5. Regression analysis

In our baseline analysis, we use OLS to estimate the regression equation

$$y_{iw} = \beta_0 + \beta_1 TG_i + \beta_2 X_i + \varepsilon_{iw}, \tag{1}$$

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 omitted 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.

4. Results

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. To this end, we focus on comparing the NoIntervention to the FactChecking and the MediaLiteracy, respectively, and defer supporting analyses of the JustFacts and the PassiveControl group to Section 6.

We consider three types of outcome: 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 the fakes are dealing with? If yes, does that affect participants' attitudes?

4.1. Intention to Treatment Effects

We start by examining the Intention to Treatment Effects (ITT) of our interventions. 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 A.1 to A.3 illustrate the results for each outcome, treatment, and wave of the survey; further details are presented below.

4.1.1. Credibility

Table 1 presents the regression results for the perceived credibility of fakes. Panel A shows the estimates from comparing the FactChecking, and Panel B from comparing the MediaLiteracy to the NoIntervention group, respectively.

Our first main result is that the fact checking intervention reduces the credibility of fakes on Corona vaccines in Wave I of the survey, but has no impact beyond that. In Wave I, participants from the FactChecking group are on average 7 to 8 percentage points less likely to perceive fakes on Corona vaccines as *Very credible* or *Credible* than participants from the NoIntervention group. The estimate is statistically significant at the 1%-level, the effect size corresponds to about 16% of a standard deviation in the dependent variable and to about 27% of its baseline 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 as *Very credible* or *Credible* than participants from the NoIntervention group. The estimate is statistically significant at the 1%-level; the effect size corresponds to about 24% of a standard deviation in the dependent variable and to 42% of its baseline value, whereby the effect is even larger than for the FactChecking group.[20] 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 17% of a standard deviation in the dependent variable and to about 7% of its baseline value in both waves of the survey.[21]

Crucially, neither intervention reduces the credibility of facts (see Table A.3 in Appendix B.2), 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 information that they encounter online.

4.1.2. Factual knowledge

Table 2 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.[22] 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

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

[21] Note that the perceived credibility of fakes on nutrition is much larger than for fakes on Corona vaccines. We further discuss this heterogeneity in Section **??**.

[22] 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 2 to be unequal to zero and one, respectively.

Table 1: 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
Mean DV	0.260	0.260	0.894	0.894	0.398	0.398	0.818	0.818
Std.Dev. DV	0.439	0.439	0.308	0.308	0.489	0.489	0.386	0.386
N	1,221	1,221	1,223	1,223	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
Mean DV	0.248	0.248	0.872	0.872	0.376	0.376	0.795	0.795
Std.Dev. DV	0.432	0.432	0.334	0.334	0.485	0.485	0.404	0.404
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020

Notes: Table 1 shows the OLS estimates of a linear probability model that compares 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* 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.

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 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 media literacy is thus smaller than the impact of fact checking when the FactChecking group receives correct information in addition to the fakes, and roughly equivalent when it does not.[23] 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 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., Nyhan, 2021; Maertens et al., 2021), which further reduces the effect size in Wave II of the survey.

Similar to the results on credibility, Table A.4 in Appendix B.2 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.

4.1.3. Attitudes

Table 3 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.

[23] 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 2: Factual knowledge 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
Mean DV	0.238	0.238	0.092	0.092	0.310	0.310	0.154	0.154
Std.Dev. DV	0.845	0.845	0.619	0.619	0.857	0.857	0.608	0.608
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
Mean DV	0.289	0.289	0.093	0.093	0.277	0.277	0.130	0.130
Std.Dev. DV	0.843	0.843	0.643	0.643	0.845	0.845	0.606	0.606
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020

Notes: Table 2 compares factual knowledge 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 mean 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.

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 *Likely* or *Very likely* to get vaccinated or 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). 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.[24] 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 more effective in swaying participants' attitudes than fact checking. In particular, participants from the MediaLiteracy group are 3.4 to 4.8 percentage points more likely to state that they are *Likely* or *Very likely* 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.[25] The effect size corresponds to about 8.7% of a standard deviation in the dependent variable and 4.2% of its baseline value in Wave I of the survey, and to about 11.5% of a standard deviation in the dependent variable and 6.2% 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 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, 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 media literacy could affect participants' attitudes, further effects from attitudes to habit change are typically modest (Verplanken and Orbell, 2022).

In sum, the results from Section 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.

[24] 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.

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

Table 3: 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
Mean DV	0.807	0.807	0.807	0.556	0.556	0.769	0.769	0.769	0.560	0.560
Std.Dev. DV	0.395	0.395	0.395	0.497	0.497	0.422	0.422	0.422	0.497	0.497
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
Mean DV	0.810	0.810	0.810	0.536	0.536	0.773	0.773	0.773	0.543	0.543
Std.Dev. DV	0.392	0.392	0.392	0.499	0.499	0.419	0.419	0.419	0.498	0.498
N	1,231	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020	1,020

Notes: Table 3 presents the OLS estimates of a linear probability model that compares 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 *Likely* or *Very likely* to get vaccinated or boostered against Covid-19, or *Unlikely* or *Very unlikely* 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.

4.2. Heterogeneity in baseline beliefs

The average impact of our fact checking and media literacy interventions is likely to depend on participants' baseline beliefs. In particular, participants who are well informed from the beginning are less likely to benefit from the interventions than participants whose baseline beliefs are poor. 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).[26] 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' baseline 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 A.5 and A.6 in Appendix B.2.[27]

4.2.1. Fact checking

Table 4 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. Second, there is ample heterogeneity in the baselines: AfD supporters are on average almost twice as likely to perceive fakes on Corona vaccines as *Very credible* or *Credible* 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, 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 4.3, where we compute persuasion rates for each of our interventions.

Note 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 (column 3). The effect is statistically significant at the 5%-level; it corresponds to 27.3% of a standard deviation in the dependent variable and to 28.7% of its baseline value. This result is especially remarkable given that fact checking typically fails to affect participants' attitudes (see Section 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 (47.8% vs. 84.1% for the non-AfD supporters), is potent enough to sway attitudes of AfD supporters who can still be convinced, while attitudes of non-AfD supporters remain unaffected.

[26] Recall that we inquired participants' political party preferences at the very end of the survey to avoid priming effects (see Section 3.1).

[27] In contrast to participants' baseline beliefs on Corona vaccines, 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 far less pronounced, patterns as for AfD supporters; see Section 6.5 for a detailed analysis.

Table 4: Heterogeneity in baseline beliefs on Corona vaccination – FactChecking

Panel A: Fact checking – AfD supporters						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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 +
Mean DV	0.404	0.329	0.478	0.667	0.261	0.455
Std.Dev. DV	0.493	0.874	0.502	0.474	0.846	0.500
N	114	115	115	99	99	99
Panel B: Fact checking – non-AfD supporters						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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 +
Mean DV	0.245	0.229	0.841	0.369	0.315	0.803
Std.Dev. DV	0.430	0.842	0.366	0.483	0.858	0.398
N	1,107	1,110	1,110	923	923	923

Notes: Table 4 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Likely* or *Very likely* 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.

4.2.2. Media literacy

Analogous to Table 4, Table 5 displays the average impact of media literacy for AfD (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 than for non-AfD supporters, and their baselines differ substantially. Hence, part of the effect heterogeneity can be explained by the different proportion of participants who can still update their beliefs (see Section 4.3 for further discussion).

Similar to fact checking, we find that media literacy 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 *Likely* or *Very likely* to get vaccinated or boostered against Covid-19 than AfD supporters from the NoIntervention group.[28] The effect is statistically significant at the 5%-level; it corresponds to about 29.7% of a standard deviation in the dependent variable and to 30.8% of its baseline value. In contrast to fact checking, this difference remains statistically significant in Wave II of the survey, and the effect size is comparable to Wave I. We thus conclude that media literacy is even more successful in swaying attitudes of participants who can still be convinced.

In sum, we find that both the fact checking and the media literacy intervention are more effective for AfD supporters, whose baseline beliefs on Corona vaccines are poor. This result stands in contrast to previous findings on motivated reasoning (e.g., Lewandowsky et al., 2012; Jerit and Zhao, 2020), whereby preexisting worldviews or attachments to a political party can impede efforts to debunk fake news. However, Pennycook et al. (2020, 2021) 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 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 if they can still do so.

4.3. Persuasion rates

Sections 4.1 and 4.2 reveal that many participants are willing to get vaccinated or boostered against Covid-19 irrespective of the interventions. As a result, the Intention to Treatment Effects are relatively small. To adjust our estimates for the share of participants left to be convinced, this section computes persuasion rates à la DellaVigna and Kaplan (2007), both for the full sample as well as for AfD and non-AfD supporters, respectively. To this end, we set the share of participants left to be convinced equal to the NoIntervention group's share of participants stating to be *Very unlikely*, *Unlikely*, or *Undecided* to get vaccinated or boostered. Then, we divide the ITT estimates from our preferred specifications ("yes +") in Tables 3, 4, and 5 by that share.

Table 6 shows that around 37.7% of our participants could still be persuaded in Wave I of the survey, and 39.2% in Wave II. Specifically, 74.6% of the AfD supporters and 33.8% of the non-AfD supporters could still be persuaded in Wave I of the survey, and 75.0% (35.2%) in Wave II. Based on that, the fact checking intervention could convince about 4.5% (8.6%) of all persuadable participants in Wave I (Wave II) of the survey.[29] The media literacy intervention could convince

[28] t-test FC ML

[29] Recall that the estimates are not statistically significant in Table 3, though.

Table 5: Heterogeneity in baseline beliefs on Corona vaccination – MediaLiteracy

Panel A: Media literacy – AfD supporters						
		Wave I			**Wave II**	
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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 +
Mean DV	0.431	0.341	0.483	0.703	0.382	0.465
Std.Dev. DV	0.497	0.859	0.502	0.459	0.774	0.501
N	116	116	116	101	101	101
Panel B: Media literacy – non-AfD supporters						
		Wave I			**Wave II**	
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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 +
Mean DV	0.229	0.283	0.844	0.341	0.265	0.806
Std.Dev. DV	0.420	0.841	0.363	0.474	0.852	0.395
N	1,115	1,115	1,115	919	919	919

Notes: Table 5 displays the effect heterogeneity between AfD supporters (Panel A) and non-AfD supporters (Panel B) 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Likely* or *Very likely* 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 9 ("yes +"), we also control for participants' Corona vaccination status.

9% of these participants in Wave I, and 12.2% in Wave II. The magnitudes are similar to those in comparable papers (e.g., Barrera et al., 2020, p.13).

Note that the persuasion rates for AfD supporters are considerably larger than the 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. A complementary explanation for our finding could be that it is *ceteris paribus* easier to convince the former group. 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 3.4 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.[30] Hence, the effect heterogeneity between AfD and non-AfD supporters is likely to be driven by poor baselines on the one hand, and uncertainty about those baselines on the other. Put differently, the average impact of our interventions is likely to increase in the share of participants left to be convinced as well as in the degree of uncertainty about their baselines.

Table 6: 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	0.045	0.184	0.018	0.086	0.120	0.085
Share to be persuaded	0.377	0.746	0.338	0.392	0.750	0.352
N	1,225	115	1,110	1,022	99	923
Panel B: Media literacy						
		Wave I			**Wave II**	
	Full	AfD	Non-AfD	Full	AfD	Non-AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Persuasion rate	0.090	0.200	0.062	0.122	0.181	0.120
Share to be persuaded	0.377	0.746	0.338	0.392	0.750	0.352
N	1,231	116	1,115	1,020	101	919

Notes: Table 6 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*, *Unlikely*, or *Undecided* to get vaccinated or boostered.

[30] 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.

5. Mechanisms

Section 4.1 shows that the effectiveness of fact checking tends to be limited to the fakes that are being corrected, while media literacy helps to distinguish between fakes and facts more generally, both in the short- and in the longer-run. A reasonable explanation is that the media literacy intervention raises participants' attention towards the Facebook postings 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 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 for further information than participants from the NoIntervention group; moreover, they become better in identifying untrustworthy elements in fakes and trustworthy elements in facts. Figures A.4 to A.6 illustrate our results; further details are given below.[31]

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 (1).

Consistent with the proposed mechanism, Table A.7 shows that all estimates for the FactChecking group are negative, but they are not statistically significant (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 10% of a standard deviation in the dependent variable and to about 11.6% (14.2%) of its baseline value for factual questions on Corona vaccines (nutrition).

5.2. Likes and dislikes

Next, we show that the media literacy intervention helps participants to identify untrustworthy elements in fakes and trustworthy elements in facts, whereby they can better distinguish between false and correct information that they encounter online. To this end, we consider the absolute number of dislikes on fakes and likes on facts (see Section 3.4) and compare participants from the NoIntervention to the FactChecking and to the MediaLiteracy group, respectively.

Table 7 confirms that the media literacy intervention induces participants to dislike fakes more often, 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 media literacy are positive and statistically significant at the 1%-level (Panel B). In Wave I of the survey, participants from the MediaLiteracy group dislike on average 1.2 more elements

[31] We do not find any systematic differences between the NoIntervention, the FactChecking, and the MediaLiteracy group in terms of the time spent with the fakes and facts. Hence, while the media literacy intervention enhances participants' ability to critically evaluate the Facebook postings, it does not induce them to engage with them for a longer period of time.

in fakes on Corona vaccines and 0.7 more elements in fakes on nutrition than participants from the NoIntervention group.[32] The effect size corresponds to 30.8% (41.5%) of a standard deviation in the dependent variable for fakes on Corona vaccines (fakes on nutrition) and to 31.9% (34%) of its baseline value. The intervention's impact persists in Wave II of the survey. Specifically, participants from the MediaLiteracy group dislike on average 0.9 more elements in fakes on Corona vaccines and 0.7 more elements in fakes on nutrition than participants from the NoIntervention group; this corresponds to 19.2% (26.2%) of a standard deviation in the dependent variable for fakes on Corona vaccines (fakes on nutrition) and to 18% (34%) of the baseline value.

Analogously, Table 8 confirms that the media literacy intervention induces participants to like facts more often, 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 media literacy are positive and statistically significant at the 1%-level (Panel B). In Wave I of the survey, participants from the MediaLiteracy group like on average 1 more element in facts on Corona vaccines and 0.7 more elements in facts on nutrition than participants from the NoIntervention group. These effects correspond to 28.8% (19.1%) of a standard deviation in the respective dependent variable and to 32.2% (18.6%) of its baseline value. Again, the impact of media literacy persists in Wave II of the survey. Specifically, participants from the MediaLiteracy group like on average 0.8 more elements in facts on Corona vaccines and 0.7 more elements in facts on nutrition than participants from the NoIntervention group; this corresponds to 31.1% (23.3%) of a standard deviation in the dependent variable for facts on Corona vaccines (facts on nutrition) and to 30% (20.6%) of the baseline value.

Crucially, participants from the MediaLiteracy group do not generally like and dislike more elements in the fakes and facts. In particular, the media literacy intervention does not increase the number of likes on fakes (Table A.8), and its impact on the number of dislikes on facts is small and limited to (arguably untrustworthy) emojis in facts on nutrition (Table A.9). Hence, consistent with the results from Section 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.

6. Further analyses

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 therefore 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 the ITTs in that regard, this section compares responses from the PassiveControl to the FactChecking and the MediaLiteracy group, respectively: the smaller

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

Table 7: Dislikes of fakes

Panel A: Fact checking

	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact checking	-0.023	-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
Mean DV	3.238	3.238	1.232	1.232	4.504	4.504	1.658	1.658
Std.Dev. DV	3.325	3.325	1.731	1.731	4.242	4.242	2.215	2.215
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
Mean DV	3.868	3.868	1.594	1.594	5.016	5.016	2.016	2.016
Std.Dev. DV	3.735	3.735	1.942	1.942	4.682	4.682	2.631	2.631
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026

Notes: Table 7 compares the absolute number of **dislikes on 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 8: Likes of facts

Panel A: Fact checking

	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact checking	0.075	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
Mean DV	2.644	2.644	3.245	3.245	2.252	2.252	3.050	3.050
Std.Dev. DV	2.926	2.926	3.271	3.271	2.215	2.215	2.893	2.893
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
Mean DV	3.134	3.134	3.608	3.608	2.671	2.671	3.393	3.393
Std.Dev. DV	3.468	3.468	3.522	3.522	2.576	2.576	3.048	3.048
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026

Notes: Table 8 compares the absolute number of **likes on 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.

the difference between those groups, the more effective is the respective intervention in repealing the impact of fakes. 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 A.2 and A.3 illustrate our results; further details are provided below.[33]

6.1.1. Factual knowledge

Table A.10 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 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. One explanation could be that the false information on nutrition is closer to participants' average prior than the false information on Corona vaccines. This would also be in line with our finding that the perceived credibility of fakes on nutrition is larger than for fakes on Corona vaccines (see Section 4.1.1). An alternative (though less 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.

6.1.2. Attitudes

In contrast to factual knowledge, Table A.11 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 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

[33] Recall that participants in the PassiveControl group are not asked to rate the credibility of fakes and facts (see Section 3.1).

participants' attitudes (see Table 3) 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.[34]

6.2. Comparison to JustFacts group

Section 5 demonstrates that the media literacy intervention helps participants to better distinguish between fakes and facts, while fact checking fails to 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 to the NoIntervention and the PassiveControl group, respectively. Figure A.2 illustrates our results, further details are discussed below.

Table XX 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.[35] 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 relative to their prior (as measured by the PassiveControl group), but reduces factual knowledge on Corona vaccines (although the JustFacts still performs significantly better than the NoIntervention group).[36] Second, Section 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 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 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 line with "anchoring" or "continued influence effects" that constitute one potential drawback of fact checking, but we cannot exclude alternative explanations, either.

[34] 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 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.

[35] 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.

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

6.3. Heterogeneity of fakes

The effectiveness of our interventions is likely to depend on the specific fakes that we select for the experiment. In particular, the more credible the fakes, the harder it is to detect them, whereby the ITTs decrease. To explore this issue in more detail, Figure A.7 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. This is roughly consistent with the results from Section 4; in particular, it could explain why the impact of our interventions tends to be larger for fakes on Corona vaccines than for fakes on nutrition.

6.4. Heterogeneity 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 A.8 displays the (disaggregated) mean credibility of all fact checks on a 5-point Likert Scale.[37]

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 facts 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 A.8 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.

6.5. 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 a play a role, we present evidence from list experiments that

[37] 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.

support the validity of the participants' self-reported attitudes, and we demonstrate that our results are robust to using an IV approach.

6.5.1. 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)– 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 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.[38]

Consistent with the robustness checks that we already conducted, Tables A.15 and A.16 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 A.15), 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 A.16 reveals that 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 4.2 and could either stem from differences in the baseline (i.e., there are more non-vaccinated participants left to be convinced) or 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.

6.5.2. Time span between Waves I and II

As we describe in Section 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 time span between actual participation in the two waves is quite heterogeneous and lies between 8 and 45 days.

We conduct two analyses to confirm that heterogeneity in the time span 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.[39] Hence, our treatments do not influence when participants re-start the survey.

[38] 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.

[39] 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.

Second, there is no evidence that differences in the time span 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 (1), all interaction terms are close to zero and statistically insignificant. Similarly, when we split the sample at the median time span (= 15 days) and estimate equation (1) on the two subsamples, respectively, the resulting point estimates resemble those from Section 4.1.[40]

6.5.3. List experiments

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. As argued in Section 3, we minimize this risk by not incentivizing the corresponding questions with a potential bonus payment. In addition, we conduct 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 C.4 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.

Table 9 shows 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, though (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.[41]

[40] Note that the time between actual participation in Wave I and II is endogenous, so all robustness checks from this section should be interpreted with caution.

[41] We further support the analysis with a sample split (see Table A.13 in Appendix B.2), 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 9: List experiments

Panel: All participants				
	Wave I		**Wave II**	
	Vacc.	Suppl.	Vacc.	Suppl.
	(1)	(2)	(3)	(4)
Direct question	18.72	30.09	21.85	30.89
List experiment	16.19	37.38	7.91	32.94
N	3,051	3,051	2,525	2,525

Notes: In row 1, Table 9 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 9 displays the respective indirectly elicited proportions from the list experiments.

6.5.4. 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 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.[42] Equation (1) thus extends to

$$y_{iw} = \gamma_0 + \gamma_1 \widehat{D_i} + \gamma_2 X_i + \epsilon_i \tag{2}$$

$$D_i = \pi_0 + \pi_1 TG_i + \pi_2 X_i + u_i, \tag{3}$$

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

[42] 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 the RA survey, the IV estimates become larger, but are qualitatively unaffected.

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.[43]

Table A.14 confirms that the 2SLS estimates of equations (2) and (3) are larger, but qualitatively similar to their counterparts from Section 4.1.[44] Moreover, the coefficients for π_1 demonstrate that close to 70% of the FactChecking, and 74% of the MediaLiteracy group spent a considerable 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.

7. Conclusion

We conduct a large-scale randomized survey experiment on the short- and longer-term effects of fact checking and media literacy interventions to demonstrate that the impact of fact checking tends to be limited to the fake news that are corrected, whereas the media literacy intervention helps to distinguish between fakes and facts more generally, both in the short- and in the longer-run. A plausible mechanism for this result is that media literacy enables participants to critically evaluate social media postings, while fact checking turns them into passive recipients of the specific corrections and thus fails to enhance their skills. Hence, in an environment where not every claim can be fact checked, media literacy 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[45], and the UNESCO has provided policy guidelines for digital media and information literacy.[46] 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.

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. E.g., some fakes are harder to detect than others, which is likely to reduce the effectiveness of our interventions. Similarly, users may be more or less well informed about different topics, whereby they are more or less likely to benefit from our interventions. Therefore, we consider the qualitative results as our most insightful findings and recommend to interpret the precise point estimates with caution.

[43] Robustness check with continuous treatment indicator.

[44] 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 A.14 in both cases, but qualitatively unaffected.

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

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

Second, 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 implementation of such trainings. In particular, it is unclear if social media platforms 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.

Third and relatedly, it is unclear how many – and especially 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 training 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 should be preferred over not doing anything.

A. Hypotheses

We pre-registered the following hypotheses in the AEA Registry under registry number AEARCTR-0008199:

Hypothesis 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.*

Hypothesis 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.*

Hypothesis 2a: *In the short- and in the longer-run, the media literacy intervention reduces the credibility of and increases factual knowledge about all "fake news" as compared to participants without intervention.*

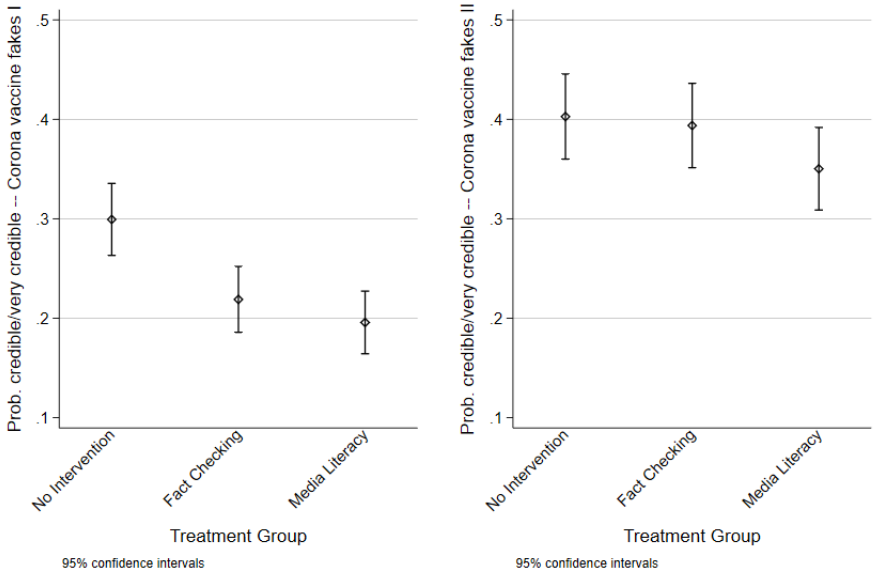
Hypothesis 2b: *In the short- and in the longer-run, participants who received the media literacy intervention are more likely to state that they are willing to get vaccinated against Covid-19 and abstain from unnecessary dietary supplements than participants without intervention.*

Hypothesis 2c: *The longer-term effects of the media literacy intervention are smaller than its short-term effects.*

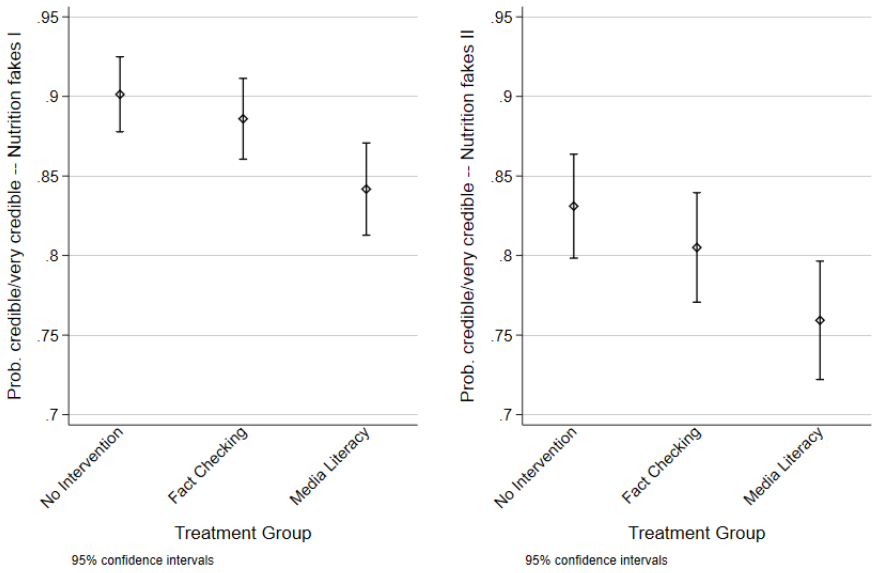
B. Omitted tables and figures

B.1. Omitted figures

B.2. Omitted tables

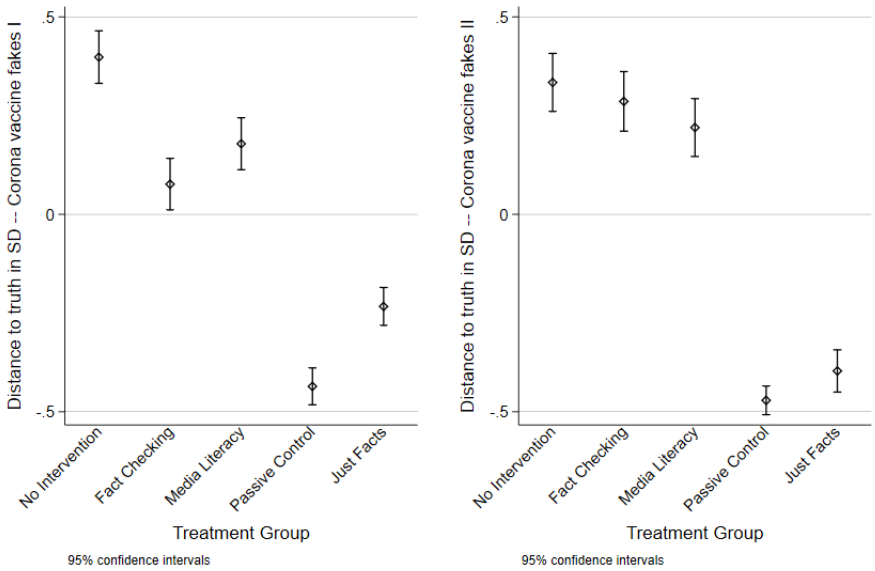


(a) Average credibility of **Corona fakes** per treatment group and survey wave

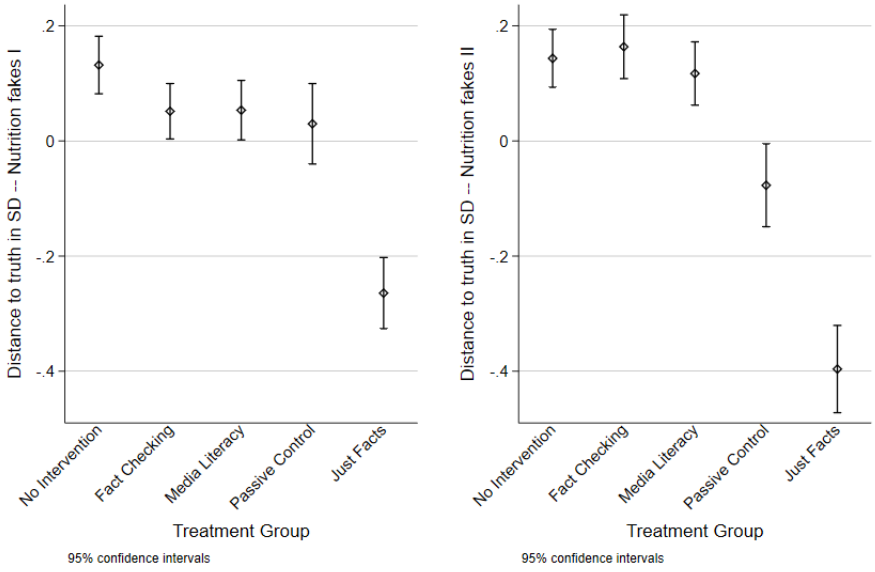


(b) Average credibility of **nutrition fakes** per treatment group and survey wave

Figure A.1: Average credibility of fakes.

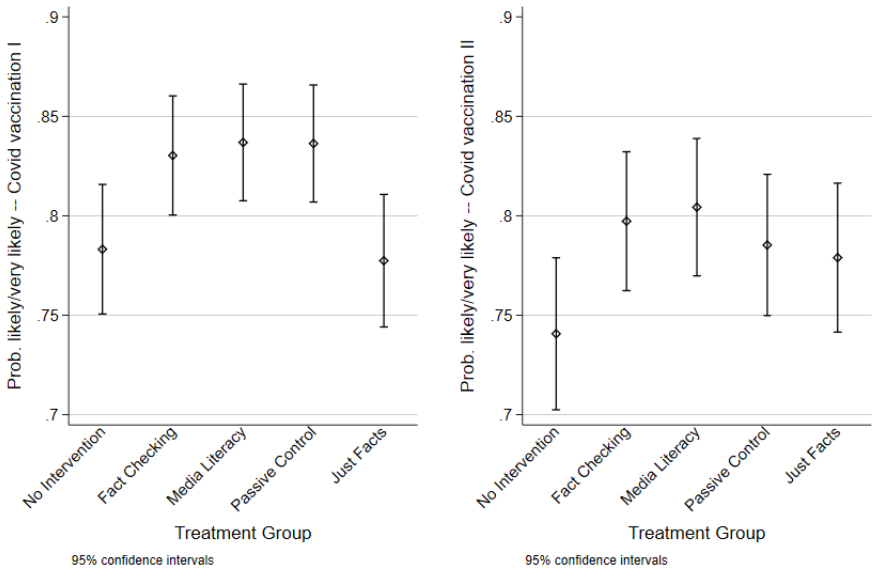


(a) Average standardized distance to the correct answers to questions on **Corona fakes** per treatment group and survey wave.

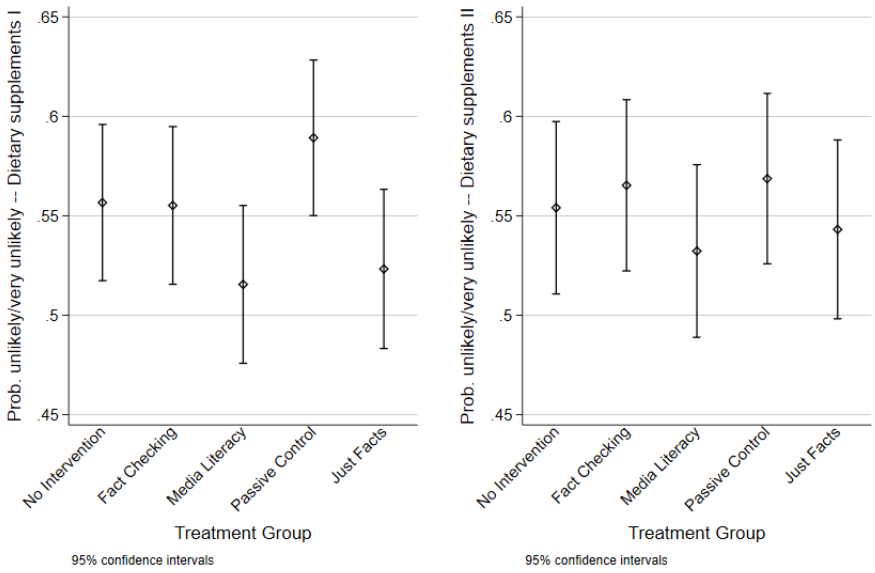


(b) Average standardized distance to the correct answers to questions on **nutrition fakes** per treatment group and survey wave.

Figure A.2: Average standardized distance to the correct answers to the factual knowledge questions on fakes.

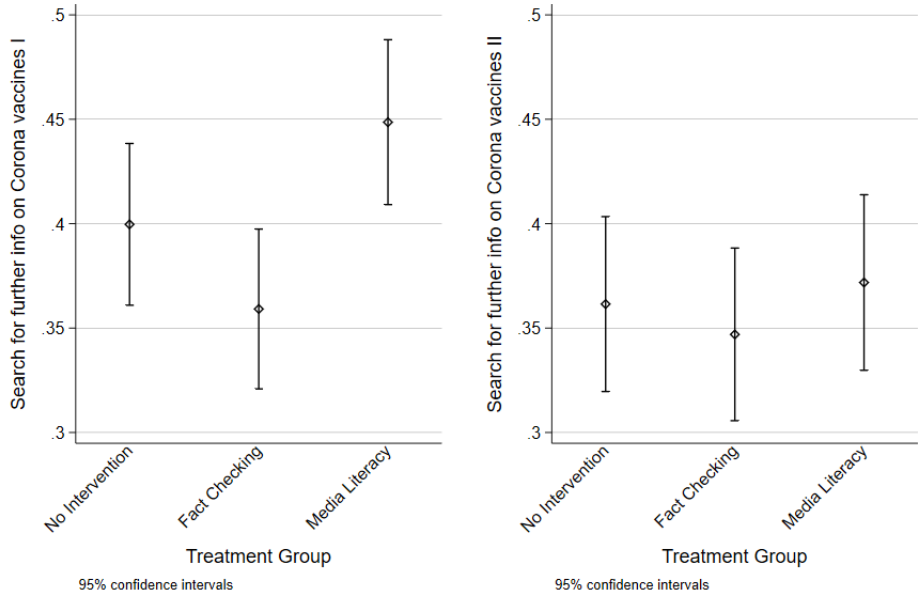


(a) Average reported probability to get vaccinated or boostered against Corona per treatment group and survey wave.

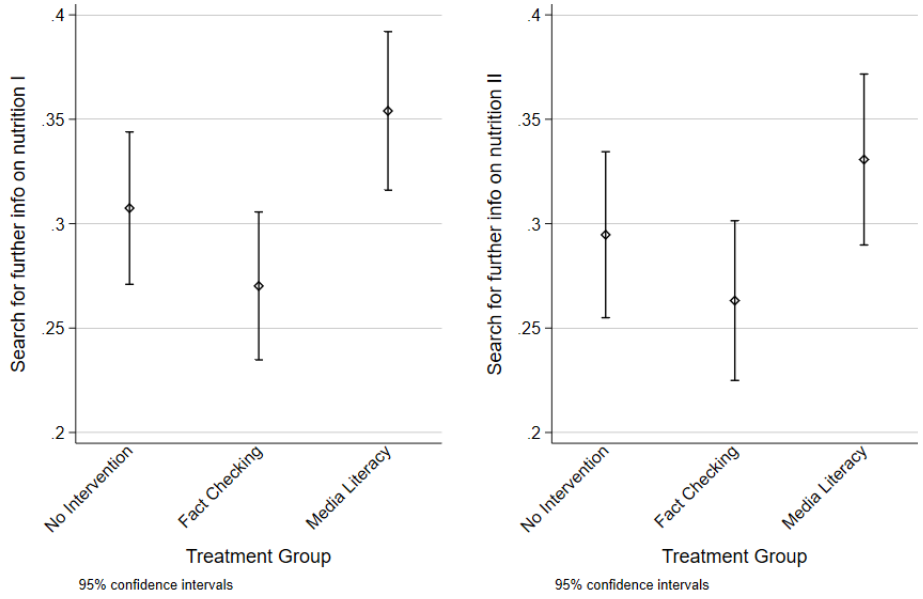


(b) Average reported probability to consume dietary supplements per treatment group and survey wave.

Figure A.3: Average reported probability to get vaccinated or boostered against Covid-19 and to consume dietary supplements.

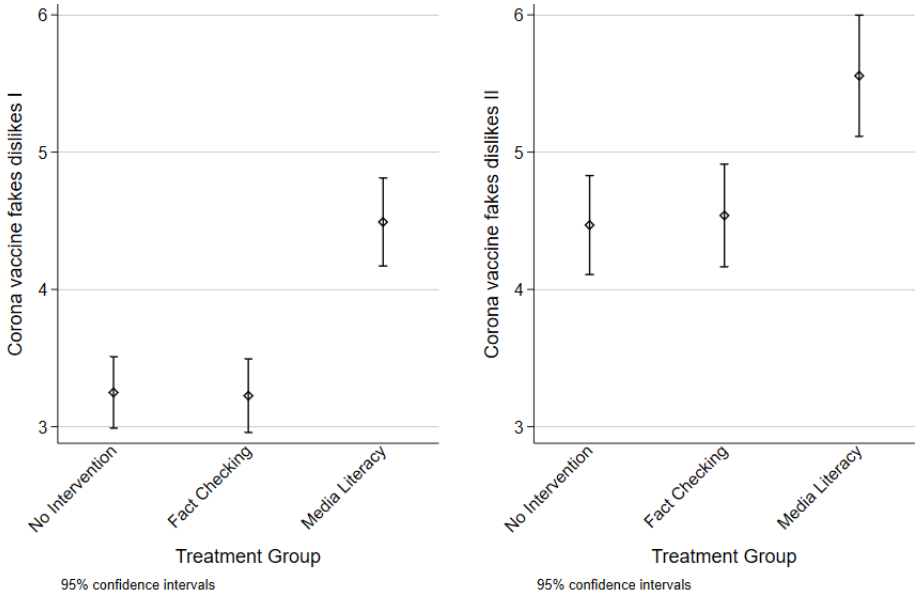


(a) Share of participants who searched for further information on Corona vaccines.

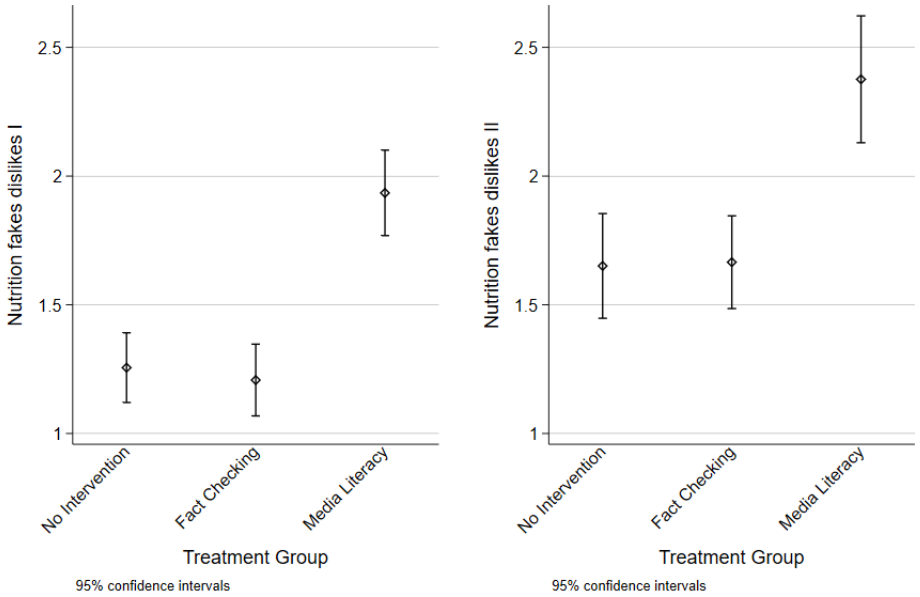


(b) Share of participants who searched for further information on nutrition.

Figure A.4: Share of participants who searched for further information online.

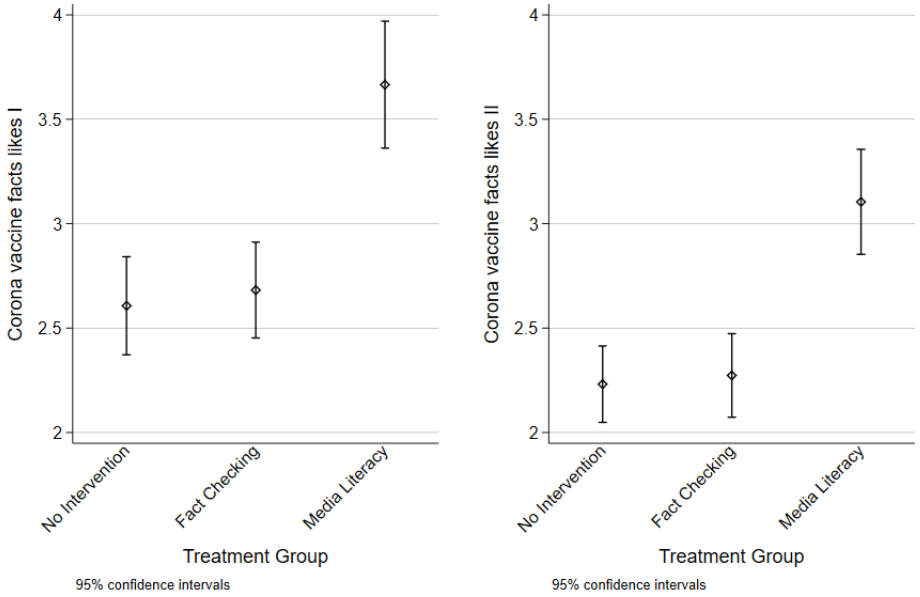


(a) Average number of dislikes for fakes on Corona vaccines.

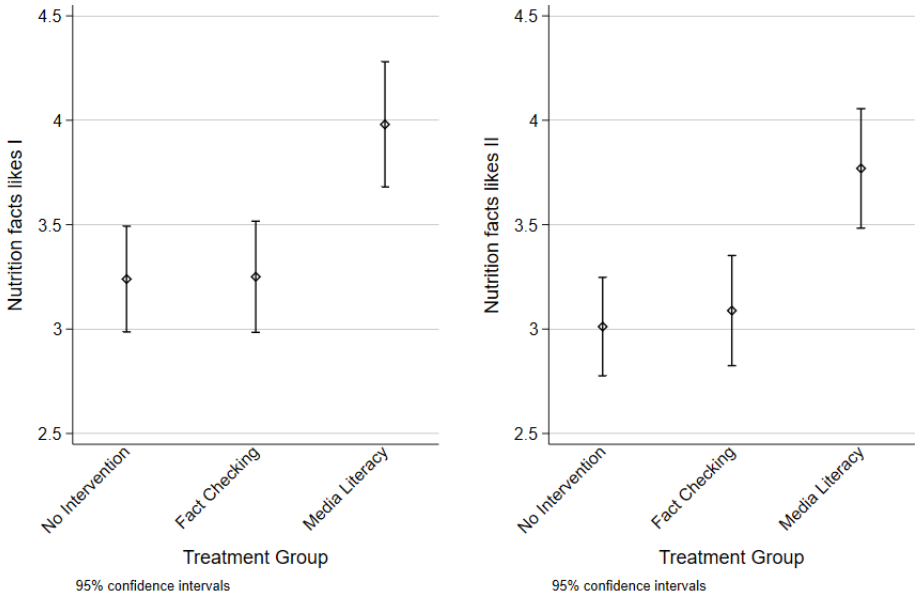


(b) Average number of dislikes for fakes on nutrition.

Figure A.5: Average number of dislikes for fakes.

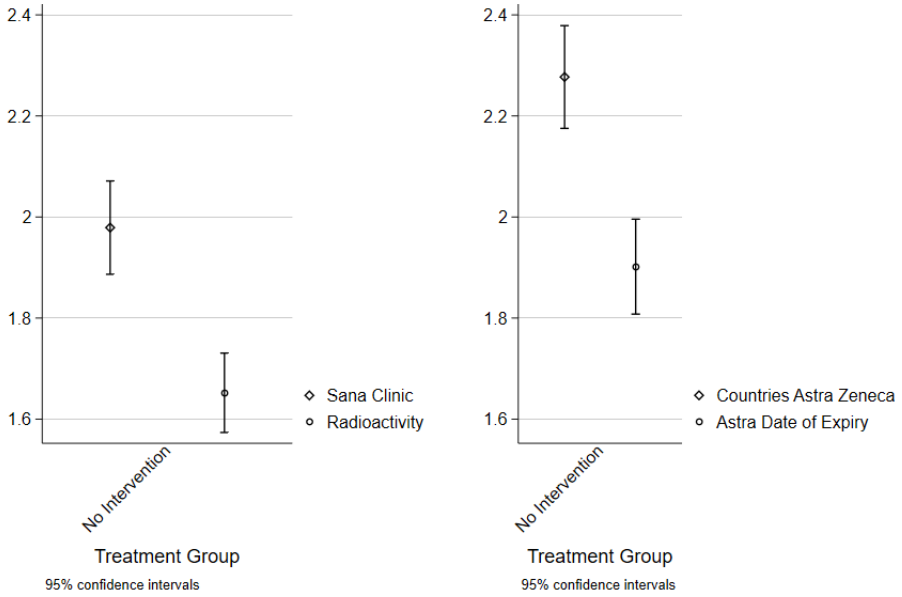


(a) Average number of likes for facts on Corona vaccines.

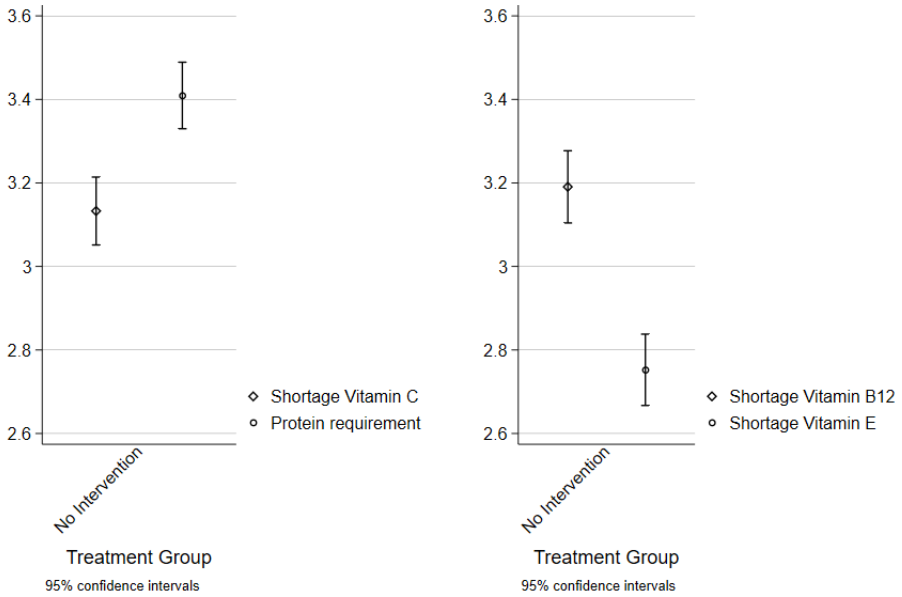


(b) Average number of likes for facts on nutrition.

Figure A.6: Average number of likes for facts.

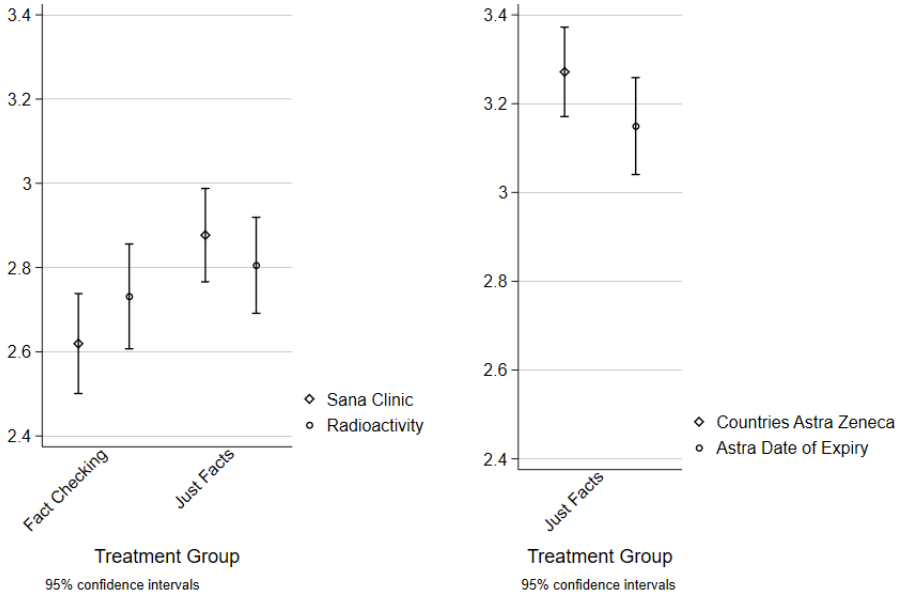


(a) Average credibility of fakes on Corona vaccines.

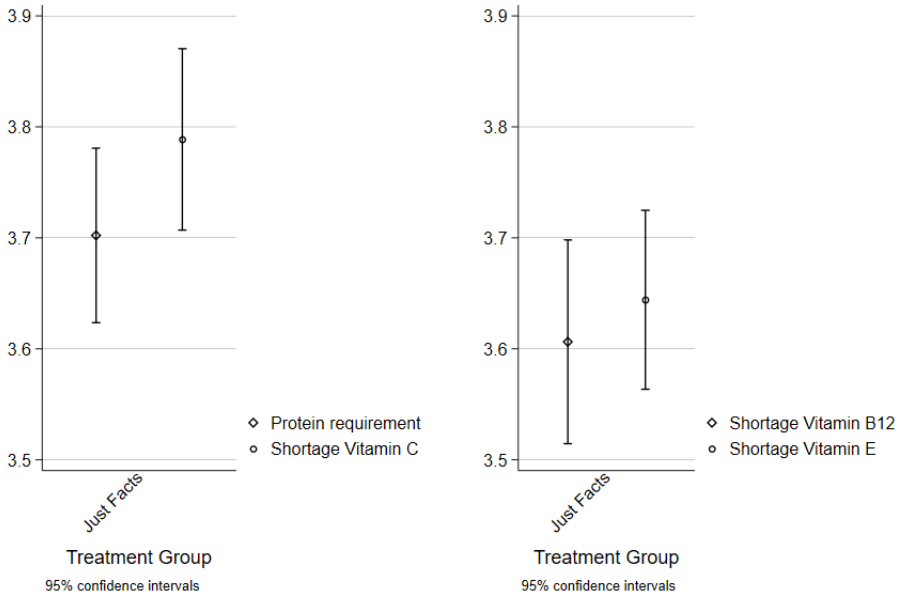


(b) Average credibility of fakes on nutrition.

Figure A.7: 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 A.8: Average credibility of fact checks on a 5-point Likert Scale.

Table A.1: 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 A.1 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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

C. Supplementary material

C.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.

C.2. Fakes, facts, and fact checks

C.3. (Un-)trustworthy areas in fakes and facts

C.4. 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

Table A.2: Summary statistics of our dependent variables

Variable	Mean	Std. Dev.	Min.	Max.	N
Credibility:					
Prob. credible/very credible – Corona vaccine fakes I	0.238	0.426	0	1	1834
Prob. credible/very credible – Nutrition fakes I	0.876	0.329	0	1	1836
Prob. credible/very credible – Corona vaccine fakes II	0.382	0.486	0	1	1533
Prob. credible/very credible – Nutrition fakes II	0.798	0.401	0	1	1533
Prob. credible/very credible – Corona vaccine facts I	0.861	0.346	0	1	2438
Prob. credible/very credible – Nutrition facts I	0.946	0.226	0	1	2439
Prob. credible/very credible – Corona vaccine facts II	0.867	0.339	0	1	2006
Prob. credible/very credible – 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. likely/very likely – Covid vaccination I	0.813	0.39	0	1	3051
Prob. unlikely/very unlikely – Dietary supplements I	0.548	0.498	0	1	3051
Prob. likely/very likely – Covid vaccination II	0.781	0.413	0	1	2525
Prob. unlikely/very unlikely – Dietary supplements II	0.553	0.497	0	1	2525
Likes and dislikes:					
Corona vaccine fakes dislikes I	3.656	3.627	0	21	1838
Nutrition fakes dislikes I	1.466	1.888	0	13	1838
Corona vaccine fakes dislikes II	4.855	4.574	0	26	1546
Nutrition fakes dislikes II	1.898	2.467	0	16	1546
Corona vaccine facts likes I	2.985	3.291	0	21	1838
Nutrition facts likes I	3.386	3.377	0	19	2440
Corona vaccine facts likes II	2.448	2.428	0	14	2028
Nutrition facts likes II	3.214	3.076	0	20	2028
Corona vaccine fakes likes I	0.671	1.419	0	17	1838
Nutrition fakes likes I	1.711	1.958	0	12	1838
Corona vaccine fakes likes I	1.42	1.998	0	15	1546
Nutrition fakes likes II	1.25	1.87	0	15	1546
Corona vaccine facts dislikes I	1.039	1.376	0	11	1838
Corona vaccine facts dislikes II	0.707	0.965	0	7	2028
Nutrition facts dislikes I	0.74	1.356	0	10	2440
Nutrition facts dislikes II	0.716	1.424	0	11	2028

Table A.3: 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
Mean DV	0.862	0.862	0.942	0.942	0.862	0.862	0.950	0.950
Std.Dev. DV	0.344	0.344	0.234	0.234	0.345	0.345	0.218	0.218
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
Mean DV	0.866	0.866	0.947	0.947	0.873	0.873	0.962	0.962
Std.Dev. DV	0.341	0.341	0.224	0.224	0.333	0.333	0.192	0.192
N	1,231	1,231	1,231	1,231	1,018	1,018	1,020	1,020

Notes: Table A.3 presents the OLS coefficients of a linear probability model that compares 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 *Credible* or *Very credible* 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 A.4: Factual knowledge 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
Mean DV	-0.064	-0.064	-0.107	-0.107	-0.152	-0.152	-0.207	-0.207
Std.Dev. DV	0.792	0.792	0.740	0.740	0.657	0.657	0.719	0.719
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
Mean DV	-0.089	-0.089	-0.089	-0.089	-0.164	-0.164	-0.208	-0.208
Std.Dev. DV	0.768	0.768	0.735	0.735	0.681	0.681	0.718	0.718
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020

Notes: Table A.4 presents OLS estimates for participants' factual knowledge 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 mean 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 A.5: Heterogeneity in baseline beliefs on nutrition – Fact checking

Panel A: Fact checking – AfD supporters

	Wave I			Wave II		
	Cred.	Knowl.	Suppl.	Cred.	Knowl.	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
Mean DV	0.868	0.161	0.600	0.869	0.174	0.596
Std.Dev. DV	0.340	0.675	0.492	0.339	0.581	0.493
N	114	115	115	99	99	99

Panel B: Fact checking – non-AfD supporters

	Wave I			Wave II		
	Cred.	Knowl.	Suppl.	Cred.	Knowl.	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
Mean DV	0.896	0.085	0.551	0.813	0.151	0.556
Std.Dev. DV	0.305	0.613	0.498	0.390	0.611	0.497
N	1,109	1,110	1,110	923	923	923

Notes: Table A.5 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Unlikely* or *Very unlikely* 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 A.6: Heterogeneity in baseline beliefs on nutrition – Media literacy

Panel A: Media literacy – AfD supporters						
	Wave I			**Wave II**		
	Cred.	Knowl.	Suppl.	Cred.	Knowl.	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
Mean DV	0.897	0.221	0.526	0.861	0.203	0.554
Std.Dev. DV	0.306	0.729	0.501	0.347	0.613	0.500
N	116	116	116	101	101	101
Panel B: Media literacy – non-AfD supporters						
	Wave I			**Wave II**		
	Cred.	Knowl.	Suppl.	Cred.	Knowl.	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

Notes: Table A.6 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Unlikely* or *Very unlikely* 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 A.7: 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
Mean DV	0.380	0.380	0.289	0.289	0.354	0.354	0.279	0.279
Std.Dev. DV	0.485	0.485	0.453	0.453	0.479	0.479	0.449	0.449
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
Mean DV	0.424	0.424	0.330	0.330	0.367	0.367	0.313	0.313
Std.Dev. DV	0.494	0.494	0.471	0.471	0.482	0.482	0.464	0.464
N	1,231	1,231	1,231	1,231	1,020	1,020	1,020	1,020

Notes: Table A.7 shows the OLS estimates of a linear probability model that compares 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.

Table A.8: Likes of fakes

Panel A: Fact checking

	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact checking	0.035	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
Mean DV	0.679	0.679	1.658	1.658	1.421	1.421	1.240	1.240
Std.Dev. DV	1.431	1.431	1.887	1.887	1.931	1.931	1.873	1.873
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
Mean DV	0.659	0.659	1.751	1.751	1.473	1.473	1.255	1.255
Std.Dev. DV	1.433	1.433	2.006	2.006	2.093	2.093	1.808	1.808
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026

Notes: Table A.8 compares the absolute number **likes on 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 A.9: Dislikes of facts

Panel A: Fact checking

	Wave I				**Wave II**			
	Corona		Nutrition		Corona		Nutrition	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fact checking	-0.156	-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
Mean DV	1.021	1.021	0.633	0.633	0.717	0.717	0.632	0.632
Std.Dev. DV	1.388	1.388	1.224	1.224	0.940	0.940	1.294	1.294
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
Mean DV	1.086	1.086	0.856	0.856	0.682	0.682	0.783	0.783
Std.Dev. DV	1.423	1.423	1.512	1.512	0.948	0.948	1.512	1.512
N	1,231	1,231	1,231	1,231	1,026	1,026	1,026	1,026

Notes: Table A.9 compares the absolute number of **dislikes on 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 A.10: Factual knowledge 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
Mean DV	-0.181	-0.181	0.041	0.041	-0.094	-0.094	0.043	0.043
Std.Dev. DV	0.751	0.751	0.754	0.754	0.780	0.780	0.752	0.752
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
Mean DV	-0.128	-0.128	0.042	0.042	-0.128	-0.128	0.019	0.019
Std.Dev. DV	0.782	0.782	0.773	0.773	0.748	0.748	0.746	0.746
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
Mean DV	-0.016	-0.016	0.081	0.081	-0.072	-0.072	0.032	0.032
Std.Dev. DV	0.838	0.838	0.766	0.766	0.777	0.777	0.726	0.726
N	1,229	1,229	1,229	1,229	1,026	1,026	1,026	1,026

Notes: Table A.10 presents OLS estimates for participants' factual knowledge 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 mean 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 A.11: 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
Mean DV	0.833	0.833	0.833	0.572	0.572	0.791	0.791	0.791	0.567	0.567
Std.Dev. DV	0.373	0.373	0.373	0.495	0.495	0.407	0.407	0.407	0.496	0.496
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
Mean DV	0.837	0.837	0.837	0.552	0.552	0.795	0.795	0.795	0.551	0.551
Std.Dev. DV	0.370	0.370	0.370	0.497	0.497	0.404	0.404	0.404	0.498	0.498
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
Mean DV	0.810	0.810	0.810	0.573	0.573	0.763	0.763	0.763	0.561	0.561
Std.Dev. DV	0.393	0.393	0.393	0.495	0.495	0.425	0.425	0.425	0.496	0.496
N	1,229	1,229	1,229	1,229	1,229	1,026	1,026	1,026	1,026	1,026

Notes: Table A.11 presents the OLS estimates of a linear probability model that compares 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 *Likely* or *Very likely* to get vaccinated or boostered against Covid-19, or *Unlikely* or *Very unlikely* 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 A.12: Factual knowledge 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	0.086	0.086	-0.082	-0.082	-0.019	-0.019	-0.117	-0.117
Std.Dev. DV	0.797	0.797	1.103	1.103	0.819	0.819	0.765	0.765
N	1,220	1,220	1,220	1,220	984	984	984	984

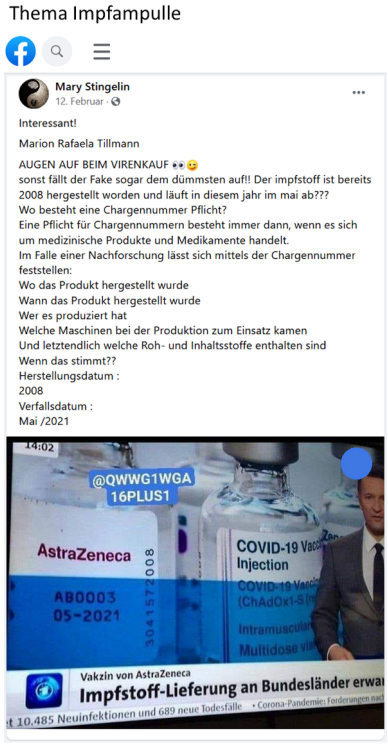
Panel B: Comparison to PassiveControl group

	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
Mean DV	-0.335	-0.335	-0.244	-0.244	-0.436	-0.436	-0.230	-0.230
Std.Dev. DV	0.603	0.603	1.203	1.203	0.511	0.511	0.851	0.851
N	1,213	1,213	1,213	1,213	992	992	992	992

Notes: Table A.12 compares factual knowledge 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 mean 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.

(a) Fake on Corona vaccines, Wave I

(b) Fake on Corona vaccines, Wave I

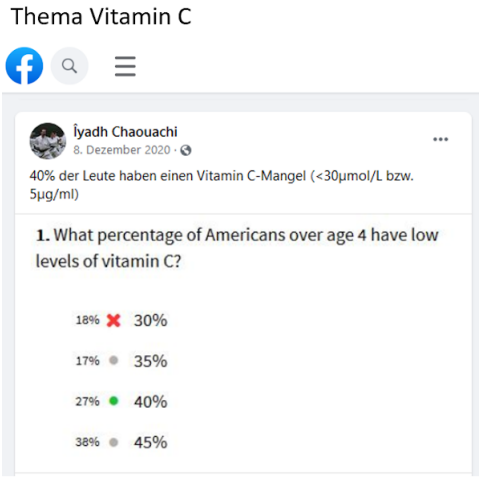


(c) Fake on Corona vaccines, Wave II

(d) Fake on Corona vaccines, Wave II

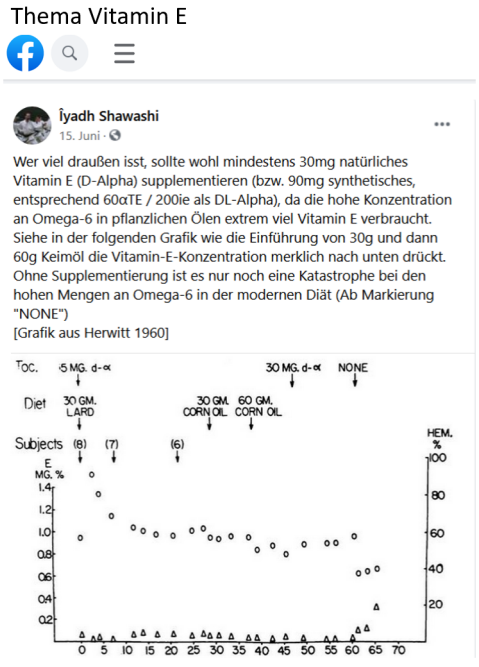
Figure A.9: Fakes on Corona vaccines.

(a) Fake on Nutrition, Wave I



(b) Fake on Nutrition, Wave I

(c) Fake on Nutrition, Wave II



(d) Fake on Nutrition, Wave II

Figure A.10: Fakes on nutrition.

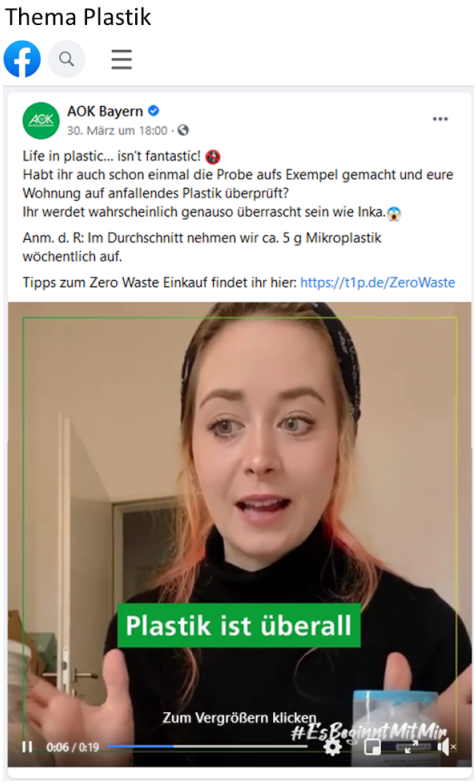
(a) Fact on Corona vaccines, Wave I

(b) Fact on Corona vaccines, Wave I

(c) Fact on Corona vaccines, Wave II

(d) Fact on Corona vaccines, Wave II

Figure A.11: Facts on Corona vaccines.



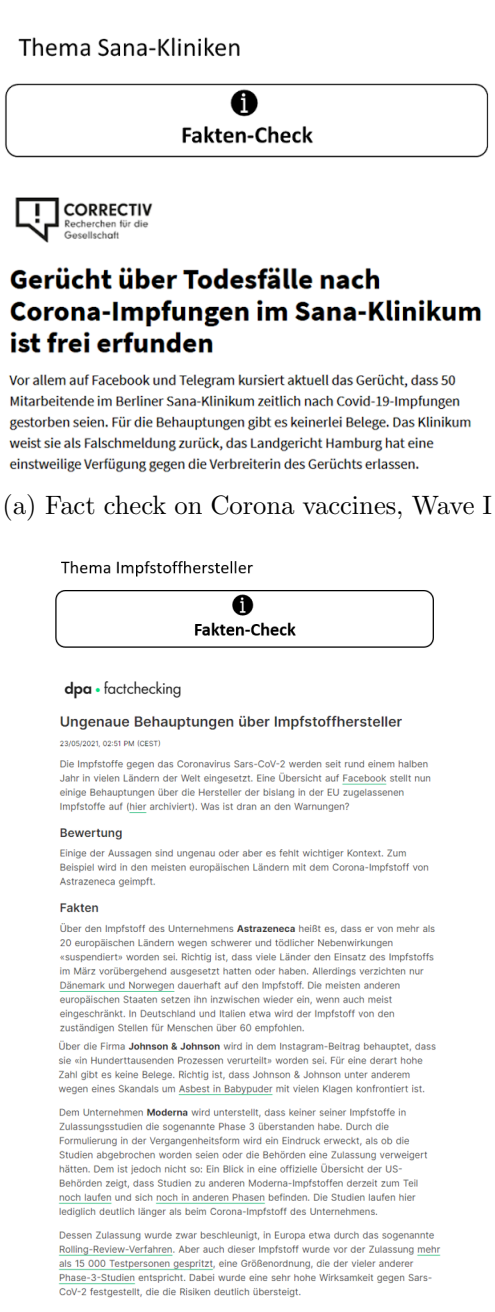
(a) Fact on Nutrition, Wave I

(b) Fact on Nutrition, Wave I

(c) Fact on Nutrition, Wave II

(d) Fact on Nutrition, Wave II

Figure A.12: Facts on Nutrition.



(a) Fact check on Corona vaccines, Wave I

(b) Fact check on Corona vaccines, Wave I

(c) Fact check on Corona vaccines, Wave II

(d) Fact check on Corona vaccines, Wave II

Figure A.13: Fact checks on Corona vaccines.

(a) Fact check on Nutrition, Wave I

(b) Fact check on Nutrition, Wave I

(c) Fact check on Nutrition, Wave II

(d) Fact check on Nutrition, Wave II

Figure A.14: Fact checks on Nutrition.

Table A.13: 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* or *Likely* 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 unlikely* or *Unlikely* to consume dietary supplements in the main experiment from those who did not and displays the respective proportions from the list experiments.

Figure A.15: Exemplary fake from the ex-post survey, where three elements are marked as “trustworthy” (in orange) and two elements are marked as “untrustworthy” (in blue).

Table A.14: 2SLS estimates for our main specifications

	Wave I						Wave II					
	Corona			Nutrition			Corona			Nutrition		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	Suppl.	Cred.	Knowl.	Vaccine	Cred.	Knowl.	Suppl.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Fact checking												
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
Mean DV	0.260	0.238	0.807	0.894	0.092	0.556	0.398	0.310	0.769	0.818	0.154	0.560
Std.Dev. DV	0.439	0.845	0.395	0.308	0.619	0.497	0.489	0.857	0.422	0.386	0.608	0.497
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.	Knowl.	Vaccine	Cred.	Knowl.	Suppl.	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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												
Fact checking	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
Mean DV	0.248	0.289	0.810	0.872	0.093	0.536	0.376	0.277	0.773	0.795	0.130	0.543
Std.Dev. DV	0.432	0.843	0.392	0.334	0.643	0.499	0.485	0.845	0.419	0.404	0.606	0.498
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

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 A.15: Heterogeneity in vaccination status – FactChecking

Panel A: Fact checking – Fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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
Mean DV	0.215	0.213	0.910	0.340	0.316	0.874
Std.Dev. DV	0.411	0.846	0.286	0.474	0.856	0.332
N	991	994	994	835	835	835
Panel B: Fact checking – Not fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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
Mean DV	0.452	0.347	0.359	0.658	0.281	0.299
Std.Dev. DV	0.499	0.834	0.481	0.476	0.862	0.459
N	230	231	231	187	187	187

Notes: Table A.15 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Likely* or *Very likely* 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 A.16: Heterogeneity in vaccination status – MediaLiteracy

Panel A: Media literacy – Fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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
Mean DV	0.200	0.274	0.905	0.315	0.259	0.874
Std.Dev. DV	0.400	0.845	0.293	0.465	0.845	0.332
N	993	993	993	835	835	835
Panel B: Media literacy – Not fully vaccinated						
	Wave I			**Wave II**		
	Cred.	Knowl.	Vaccine	Cred.	Knowl.	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
Mean DV	0.450	0.350	0.412	0.654	0.357	0.314
Std.Dev. DV	0.498	0.833	0.493	0.477	0.840	0.465
N	238	238	238	185	185	185

Notes: Table A.16 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* or *Credible* on average. In columns 2 and 5, the dependent variable is equal to participant i's mean 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 *Likely* or *Very likely* 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.

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.**

References

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva (2022) "Immigration and Redistribution," *The Review of Economic Studies*.

Allcott, Hunt and Matthew Gentzkow (2017) "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, **31** (2), 211–36.

Bailey, Regan L, Jaime J Gahche, Paige E Miller, Paul R Thomas, and Johanna T Dwyer (2013) "Why US adults use dietary supplements," *JAMA Internal Medicine*, **173** (5), 355–361.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya (2020) "Facts, alternative facts, and fact checking in times of post-truth politics," *Journal of Public Economics*, **182**, 104123.

Blair, Graeme and Kosuke Imai (2012) "Statistical analysis of list experiments," *Political Analysis*, **20** (1), 47–77.

Bode, Leticia and Emily K Vraga (2015) "In related news, that was wrong: The correction of misinformation through related stories functionality in social media," *Journal of Communication*, **65** (4), 619–638.

Chiou, Wen-Bin, Chao-Chin Yang, and Chin-Sheng Wan (2011) "Ironic effects of dietary supplementation: illusory invulnerability created by taking dietary supplements licenses health-risk behaviors," *Psychological Science*, **22** (8), 1081–1086.

DellaVigna, Stefano and Matthew Gentzkow (2010) "Persuasion: Empirical Evidence," *Annual Review of Economics*, **2** (1), 643–669.

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

Gensing, Patrick (2021) "Wie die AfD Angst vor Impfungen schürt," *Tagesschau.de*.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer (2019) "Fake news on Twitter during the 2016 US presidential election," *Science*, **363** (6425), 374–378.

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar (2020) "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India," *Proceedings of the National Academy of Sciences*, **117** (27), 15536–15545.

Guess, Andrew, Jonathan Nagler, and Joshua Tucker (2019) "Less than you think: Prevalence and predictors of fake news dissemination on Facebook," *Science Advances*, **5** (1), eaau4586.

Guess, Andrew, Brendan Nyhan, and Jason Reifler (2018) "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign," *European Research Council*, **9** (3), 4.

Henry, Emeric, Ekaterina V Zhuravskaya, SM Guriev et al. (2020) *Checking and sharing alt-facts*: Centre for Economic Policy Research.

Hopkins, Daniel J (2009) "No more Wilder effect, never a Whitman effect: When and why polls mislead about black and female candidates," *The Journal of Politics*, **71** (3), 769–781.

Jerit, Jennifer and Yangzi Zhao (2020) "Political misinformation," *Annual Review of Political Science*, **23**, 77–94.

Kuklinski, James H, Paul J Quirk, Jennifer Jerit, David Schwieder, and Robert F Rich (2000) "Misinformation and the currency of democratic citizenship," *The Journal of Politics*, **62** (3), 790–816.

Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, and David Rothschild (2018) "The science of fake news," *Science*, **359** (6380), 1094–1096.

Lewandowsky, Stephan, Ullrich KH Ecker, Colleen M Seifert, Norbert Schwarz, and John Cook (2012) "Misinformation and its correction: Continued influence and successful debiasing," *Psychological science in the public interest*, **13** (3), 106–131.

Luca, Michael (2015) "User-generated content and social media," in *Handbook of Media Economics*, **1**: Elsevier, 563–592.

Maertens, Rakoen, Jon Roozenbeek, Melisa Basol, and Sander van der Linden (2021) "Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments.," *Journal of Experimental Psychology: Applied*, **27** (1), 1.

Nyhan, Brendan (2021) "Why the backfire effect does not explain the durability of political misperceptions," *Proceedings of the National Academy of Sciences*, **118** (15), e1912440117.

Nyhan, Brendan, Ethan Porter, Jason Reifler, and Thomas J Wood (2020) "Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability," *Political Behavior*, **42** (3), 939–960.

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

——— (2015) "Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information," *Vaccine*, **33** (3), 459–464.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand (2021) "Shifting attention to accuracy can reduce misinformation online," *Nature*, **592** (7855), 590–595.

Pennycook, Gordon, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand (2020) "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention," *Psychological Science*, **31** (7), 770–780.

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

——— (2021) "The Psychology of Fake News," *Trends in Cognitive Sciences*, **25** (5), 388–402.

Radimer, Kathy, Bernadette Bindewald, Jeffery Hughes, Bethene Ervin, Christine Swanson, and Mary Frances Picciano (2004) "Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000," *American Journal of Epidemiology*, **160** (4), 339–349.

Roozenbeek, Jon, Sander van der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky (2022) "Psychological inoculation improves resilience against misinformation on social media," *Science Advances*, **8** (34), eabo6254, URL: `https://www.science.org/doi/abs/10.1126/sciadv.abo6254`.

Stantcheva, Stefanie (2021) "Understanding tax policy: How do people reason?" *The Quarterly Journal of Economics*, **136** (4), 2309–2369.

Swire, Briony, Adam J Berinsky, Stephan Lewandowsky, and Ullrich KH Ecker (2017) "Processing political misinformation: comprehending the Trump phenomenon," *Royal Society Open Science*, **4** (3), 160802.

Tversky, Amos and Daniel Kahneman (1974) "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.," *Science*, **185** (4157), 1124–1131.

Verplanken, Bas and Sheina Orbell (2022) "Attitudes, habits, and behavior change.," *Annual Review of Psychology*.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018) "The spread of true and false news online," *Science*, **359** (6380), 1146–1151.

Vraga, Emily K and Leticia Bode (2017) "Using expert sources to correct health misinformation in social media," *Science Communication*, **39** (5), 621–645.

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