ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis Economics and Business Economics

Fighting COVID-19 misinformation with awareness-raising interventions

Name student: Stefan Cojocariu

Student ID number: 535919

Supervisor: Marine Hainguerlot

Second assessor: Diarmaid O Ceallaigh

Date final version: 31/07/2022

Table of contents

Abstract

Information is an extremely valuable resource in our modern age. Through these times where any click can cause multiple stimuli to be sent to our brains, evaluating the reliability of the information picked up is more crucial than ever. For this reason, this paper compares the use of a forewarning message and a guideline assessment message to counter the influence of the most topical misinformation we are exposed to: COVID-19 fake news. Running a random controlled trial with survey-collected data, true and fake news headlines' accuracy ratings are analyzed. While no conclusive effect of these two interventions on accuracy reports is found, relevant implications for future research on this topic are provided.

Introduction

When asked about their exposure to fake news in a survey, more than a third of 26 thousand participants throughout the European Union reported coming across misinformation daily (European Commission, 2018). While the topic of fake news started reaching more people's awareness during the 2016 Elections of the United States, the global pandemic caused by the COVID-19 let the misinformation problem reach even higher scales (Pennycook et al., 2020b). During this worldwide event, rumors, hoaxes concerning sanitary practices and even conspiracy theories about the virus' origins as a Chinese bioweapon have all contributed to the occurrence of an infodemic. According to the World Health Organization (WHO), an infodemic refers to too much information being available about an issue in the middle of an epidemic outbreak (WHO, 2020). This plethora of news hinders the distinction between real and fake information, misguiding people's trust and behavior regarding health measures and experts on the topic. In fact, Roozenbeek et al. (2020) have demonstrated that higher vulnerability to false information decreases not only conformity to COVID-19-related health instructions and policies, but also individuals' inclination to get the vaccine and encourage at-risk relatives of getting it as well. It can be established that the need for reliability discernment for news has become a crucial, day-to-day challenge. Therefore, interventions targeting such misinformation and impeding their spread, by means of factual and preventive messages for instance, are worth researching. This paper will be focused on testing the effect that exposure to distinctive awareness-raising methods can have on information discernment. Thus, the research question is formulated as follows:

How can exposure to different awareness-raising methods influence COVID-19 fake news perception?

To answer this inquiry, data on the perception of different COVID-19 news will be collected via a survey. Respondents will be divided into 2 groups, each exposed to an awareness-raising stimuli and then asked to evaluate its usefulness. Data from both groups will then be assessed and compared, in hopes of identifying an intervention, or specific elements, promoting more in-depth information appraisal.

Theoretical framework

Previous literature on means of fighting fake news has explored psychological models such as the theory of inoculation, which is based on behavioral cognitive arguments. Defined as a method of gaining immunity to misinformation by being exposed to a "weakened" form of the latter (Banas &

Rains, 2010), the biological analogy with the vaccine-induced immunity is fitting, given the health-related context of this study. Drawing further from the medical discipline, this approach promotes prevention, rather than cure as a means of fighting misinformation. Consisting of two key features: a forewarning and a prebunking process (scientifically referred to as refutational preemption), inoculation uses the former as a warning for a forthcoming threat and the latter as a tool to disprove a fallacious argument (Cook et al., 2017). For instance, an inoculation could include a straightforward indication that misinformation regarding vaccine's side effects is actively shared on social media, followed by a brief list of common fake symptoms associated with covid vaccination, explaining how each of them was fabricated. While the effectiveness of inoculation has been proved in numerous studies on building resistance to social pressure-induced smoking, fake news on climate change, detrimental credit card marketing targeting college students, and even more pertinently, COVID-19 misinformation (Pfau et al., 1992; van der Linden et al., 2017; Compton & Pfau, 2004; Islam et al., 2021), this approach also has its limitations.

As mentioned by Compton (2013), timing is a problematic aspect of inoculation theory since similarly to medical inoculation, human bodies require time to develop antibodies from the threat they face. In context of an experimental study, such longer timing can be extremely hard to implement as participants must already allocate time and attention to fill in survey answers and process the information presented to them. This can render inoculation techniques harder to experimentally implement and evaluate within limited time constraints. As a method to raise awareness, this study will first explore the use of the following message: "Don't believe everything you read online". The intuition behind the formulation of this sentence resides in a short, psychological reminder to think twice before taking information as given and to always stay aware by applying critical thinking. This approach is backed by multiple works of Pennycook & Rand (2019, 2021), who have observed in different environments that engaging in extensive cognitive reflection enables a better discernment of true and fake news. This awareness-raising method will be tested against an inoculation-inspired intervention where participants are asked to read and rank different guidelines on how to avoid misinformation. Not only is ranking a tool to ensure active participation and a better understanding of the message conveyed, by way of higher internal consistency (Lutzke et al., 2019), but when pursued in consistency with individual principles, ranking is a core feature of rationality (Simon, 1955). Similar interventions have already been successfully implemented by Lutzke et al., in context of climate-related fake news and by Guess et al. (2020) in India and the United States.

In fact, the two interventions studied consist of the two elements of inoculation theory: the forewarning message "Don't believe everything you read online" serves as the straightforward indication preparing a person for an imminent misinformation threat, while the guidelines rating process serves as a refutational preemption, arming the exposed individual with the media literacy tools to counter misinformation. The reasoning behind the choice of these two specific interventions thus comes from the desire to test the two characteristics of inoculation against each other, to determine whether these two interventions can, on their own, have a positive impact on fake news prevention. The idea itself of comparing these two elements comes from the abovementioned experimental limitations of inoculation denoted by Compton, combined with another, more widespread timing issue: limited attention span. Since the beginning of the COVID-19 pandemic, vast amounts of students have reported even shorter attention spans than before, a problem which is especially salient in information-heavy, instructive online environments (Mukhtar et al., 2020). Drawing a parallel between the argument-dense refutational preemption and the resource intensive remote teaching environment, support for a more concise and grabbing warning message can be considered. Consequently, the potential of using a forewarning as a tool to fight misinformation is formulated through the following hypothesis:

H1: "*Using a forewarning message helps misinformation discernment more than assessing guidelines on how to detect fake news*"

Comparing the forewarning and the guideline evaluation method could lead to new policy implications for the use of preventive messages to hinder fake news spread, such as the redesign of social media applications and websites, to include fake news prevention warnings at the top of their main page or feed. While papers defend the need for both a forewarning and a prebunking stage in inoculation theory (Pfau, 1992), none were found to contrast its two core features, while also considering the empirical limitations of inoculation, in terms of a post-lockdown online survey. Besides being the first work on COVID-19 fake news to test this specific forewarning intervention and to contrast it with the guidelines ranking intervention, this paper has other novelty claims. Previous literature has denoted that older adults are more likely to believe and share fake news than younger ones (Brashier & Schacter, 2020) and are more inclined to suffer from memory distortion after being exposed to misinformation (Wylie et al., 2014). An article by Guess, Nagler and Tucker (2019) reveals that even when education, partisanship biases and ideologies are held constant, older people, specifically above the age of 65, are more likely to share fake news on Facebook than any other age group; age being the only demographic variable found to always impact fake news spread. This indicates that not all age groups are affected in the same way by misinformation, with older people being more at risk of being misled. No other study has led an inoculation-based analysis of the effects of COVID-19 fake news discernment methods on specific age demographics. In hopes of finding conclusive results and building on existing literature, the following secondary hypothesis is formulated:

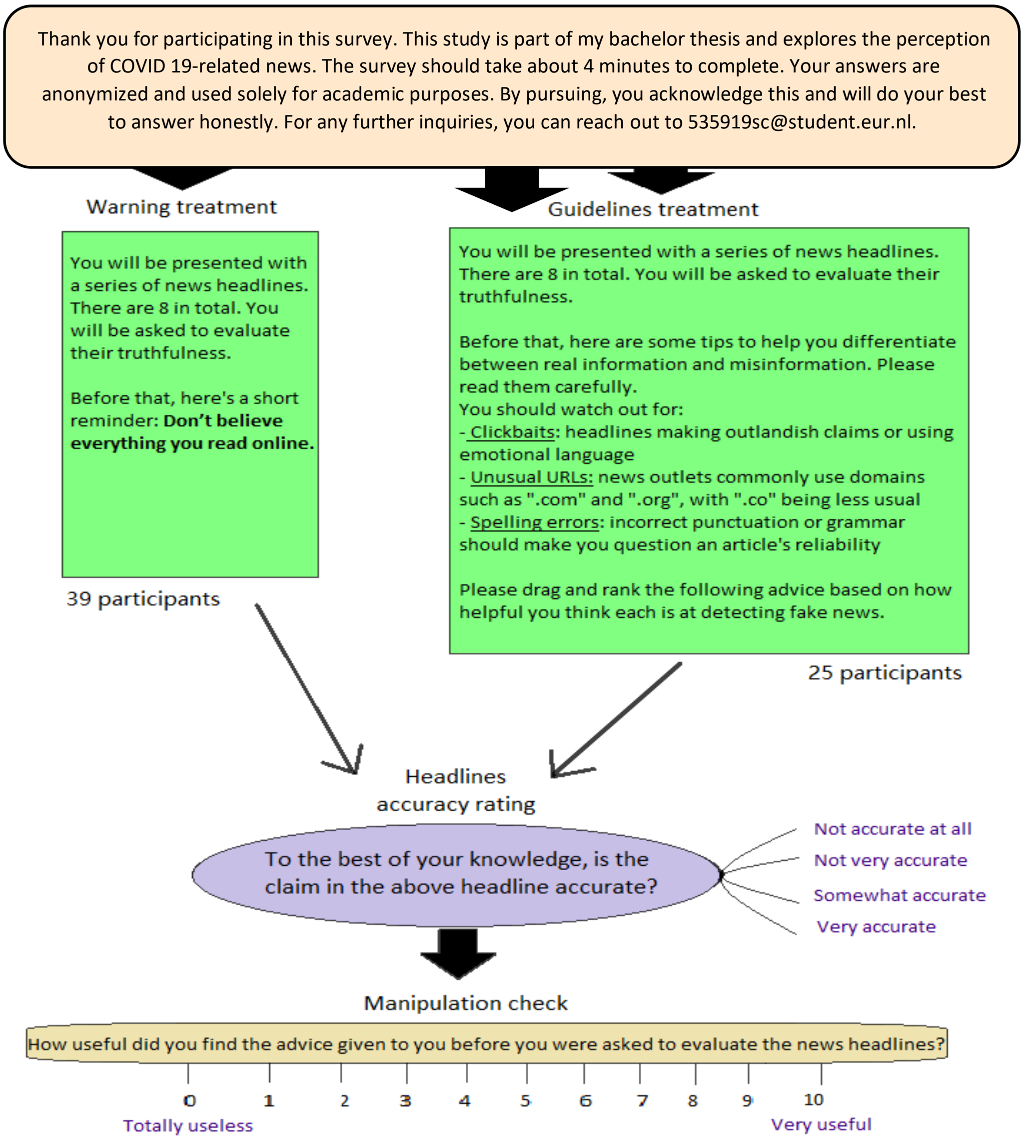
H2: *"Older adults show reduced misinformation discernment compared to younger adults, for both awareness-raising methods"*

Methods

Survey

To test the first hypothesis, an online survey has been designed through Qualtrics. To fulfill visualization purposes and facilitate the lecture of this subsection, Figure 1.1 depicts the flow of the survey conducted.

Figure 1.1. Illustration of the survey



The survey starts with a consent form, declaring the average response time needed to complete the survey, explaining the purpose of the survey, and ensuring the confidentiality of the participants' data as well as its academic usage. The respondents are informed that by pursuing the survey, they agree with the abovementioned conditions and engage themselves in answering honestly. Then, the study inquires about demographic-related information such as the age, nationality, gender, and education level of the participants. Following that, the participants are randomly divided into one of two treatment groups. The Warning group is presented with the "Don't believe everything you read online" message while the Guidelines group is presented with article-sourced tips on how to detect fake news, then asked to rank them based on their usefulness. Regarding the tips presented, they consist of warnings on clickbaits, unusual website sources and spelling mistakes. An in-depth description of these guidelines is included in Figure 1.1, beneath the "Guidelines treatment" heading. The tips are obtained from the fact-checking website snopes.com and have been selected based on how common such occurrences are actually present in the media, especially in the context of COVID-19 and more broadly health-related news (Snopes, 2022). While more extensively discussed in the *Data* section, it is worth noting that the exposure time to each treatment is tracked and then reported, notably for the data cleaning process. It is also important to mention that the methods used in this paper are based on the *practical guide to doing behavioral research on fake news and misinformation* by Pennycook et al. (2021).

Following the attribution to a treatment cluster, both groups are shown 8 images. Half of them are depicting carefully selected circulating COVID-19 fake news headlines while the other half consists of fact-checked, true news. The following sub-section *Materials* discusses the headlines used and how they are selected in more detail. To ensure that the sequencing of the news material does not influence in any way the answers provided by the participants, the order in which the news headlines are presented is randomized. For similar reasons, a page break is added between each headline to avoid respondents strategically assessing the content, or in other terms, to prevent them from first reading all the headlines on the same page and then rating them one by one. Under each headline, the fake news discernment, measured as belief in fake news subtracted from belief in true news, is addressed by the question "To the best of your knowledge, is the claim in the above headline accurate?" with answers ranging from "Not accurate at all", "Not very accurate" to "Somewhat accurate" and "Very accurate". Empirically, both fake and true news discernment is measured on a 4-point Likert scale. This response scale is mainly present in surveys where the respondents' level of agreement with a statement is sought (Bertram, 2007). Scale value 1 is assigned to the answer "Not accurate at all", 2 to "Not very accurate", 3 to "Somewhat accurate" and 4 to "Very accurate". The wording of the accuracy-evaluating question & answers as well as the

4-point Likert scale have been previously employed by Pennycook et al. (2020a, 2021) within the framework of misinformation differentiation assessment. The final part of the survey consists of a manipulation check, where the question **"**How useful did you find the advice given to you before you were asked to evaluate the news headlines?" is asked with an answer scale from 0 to 10 with 0 referring to "Totally useless" and 10 to "Very useful". The goal of this measure is to verify that both treatments have been successfully attributed to the participants, meaning that their news discernment abilities have been influenced by the exposure to the awareness-raising interventions.

Materials

Concerning the news headlines presented in the survey, 4 were collected on the fact-checking website snopes.com and the other 4 were selected from the material made available by Pennycook et al. (2020b). Snopes is a renowned website for fact checks and some of its content labelled as "junk news" is very reflective (or sometimes, literally consists of) fake news headlines that can be commonly found on Twitter or Facebook, making it very appropriate for the purpose of this paper. The collected materials were arranged in the "Facebook format", based on Pennycook et al.'s *practical guide to doing behavioral research on fake news and misinformation* (2021), to help generalizability and use a heavily recognized fake news-sharing layout. An important issue also mentioned in the guide is the use of specific awareness-raising signals and basing the headline material on these warnings rather than the common online and real-world headlines. For example, instead of selecting headlines explicitly containing spelling errors and using emotional language to favor results from the treatment recommending misinformation discernment guidelines, the selection process applied to the headlines focused on distinct factors. Such factors include having minimal time-context dependency (as part of the previously used materials were time-specific and risk being outdated) and addressing the COVID-19 misinformation subject through a medical lens, by using content employing terms such as "ivermectin", "hypoxia" and "blood-clot" which, as reflected by their increased presence on fact-checking websites, attests their relevancy within the Covid information sphere. The latter element is a very important part of the selection process, as the goal of the headline material is to reflect what is happening in the world. The news headline content is included in the Appendix section of this thesis. To better understand the selected material, some brief motivation for each headline is provided. The first true news headline in Figure A1 and the first fake news headline in Figure A5 both have the COVID-19 vaccine as key element. The former refers to Pfizer's vaccine high effectiveness rate while the latter mentions Johnson & Johnson's vaccines restrictions due to blood-clot risks. Regarding the first fake news headline, the deceit resides in the difference between the term “effectiveness”, which is incorrectly used in the title and represents the real-world functioning of the vaccine based on data reports over the years, and the term “efficacy”, which should have been the one employed as it depicts the vaccine’s performance under a supervised experimental setting. The second and third fake news contents share the common characteristic with the first true news content (illustrated respectively in Figures A2, A3 and A5) in that they all contain specific medical terms. These terms are mentioned above, and their importance can be restated through their recent prevalence in online searches and media. Regarding the remaining true news and the fourth fake news headline, they have been selected across the content made available by Pennycook et al., with the main selection criterion of minimizing time specificity. The fourth fake news article illustrated by Figure A4 is purposefully vague in its choice of words as it states that some scientists *may* have found the coronavirus cure. Similarly, the third true news content in Figure A7 might induce some confusion as it mentions a *false* claim spread by the US police regarding meth contamination with coronavirus. The presence of the term “false claim”, added to the topic of methamphetamines and the source being Buzzfeed (website known for trivial reports), can send mixed signals as to the trustworthiness of this headline.

Data

Data description

To investigate the potential of awareness-raising interventions to combat misinformation, the primary data collected from the abovementioned survey is used. The survey was mainly shared via WhatsApp to fellow students, friends, and wide-scale family. In total, 77 responses were gathered during the data collection period (which started on the 25th of May and ended on the 3rd of June). Once retrieved, the responses go through the data cleaning process. First, 2 respondents are removed due to incomplete answers and not finishing the survey. The total duration of the survey should be above 2 minutes and under 100 minutes for every participant. These limitations are set in place to sort and invalidate too hasty or inactive answers and have the outcome of eliminating 5 answers. Following that, the time spent on the page displaying the treatment message is measured and evaluated. This is done to guarantee that both groups face sufficient exposure to the treatments. For the Guidelines group, any exposure under 20 seconds is eliminated. For the Warning group, any response under 6 seconds is eliminated. These specific time values have been selected based on the shortest time needed to read and process the treatment information. Following this process, 6 participants are removed from the Guidelines group. Once these data selection precautions are set, 64 valid responses remain. Based on Qualtrics' randomization process, 39 subjects have been allocated to the Warning group and 25 to the Guidelines group. Table 1.1 below displays demographic statistics collected from the candidates.

Table 1.1. Descriptive statistics of demographics

	Mean	**Stand. Dev.**
Age	36.89	12.08
Education level		0.90
High school	0.31	
Bachelor	0.34	
Masters	0.30	
Ph. D.	0.05	
Country		0.03
Romania	0.63	
Belgium	0.08	
UK	0.08	
Germany	0.05	
Spain	0.03	
Other	0.14	
Gender		0.50
Male	0.41	
Female	0.59	
Observations	64	

The mean age of the participants is about 37 years old, and the median age is 38 years. 31% have finished high school and about 69% of the participants have completed tertiary education. The dominant nationality is Romanian, representing 63% of the subjects observed (this can be explained by my Romanian origins and the distribution of the survey to my family and their colleagues), with Belgian and British tying at second place with 8% of the respondents. About 14% of the participants individually come from different countries of the world. Regarding the gender of the respondents, more women than men have answered the survey, respectively defining 59% and 41% of those who answered the survey.

The collected demographic variables are compared and used to test the secondary hypothesis. As younger adults are defined as aged 18 to 29 and older adults as older than 65 years old in Brashier & Schacter's (2020) study, these scales and categories are revised based on the ages of the respondents. The youngest respondents are 20 years old and the oldest one is 57 years old. Since no respondent fits the accepted “older adult” age cluster which is dependent on the retirement age and usually starts as early as 60 to a more common 65 years of age, the sample is split at the median age of 38 to distinguish between younger and older adults. The former age cohort contains adults under the age of 38 while the latter contains adults over that age.

To verify whether the randomized assignment to one of the two treatments groups functioned properly, a randomization check is required. To perform the randomization check, the means of all 4 demographic variables, namely age, gender, country of origin and highest education level completed are compared across both groups by using a two-sample t test and then reported on Table 1.2.

Table 1.2. Randomization check t-test for all demographic variables

	Treatment W			**Treatment G**			**t-test**				
Variable	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**t**	**df**	**Diff.**	**p***	**Decision**
age	39	39.21	11.57	25	33.28	12.19	1.96	62	5.93	0.055	Don't Reject
gender	39	1.64	0.49	25	1.52	0.51	0.95	62	0.12	0.344	Don’t Reject
country	39	8.77	3.17	25	9.12	3.47	-0.42	62	-0.35	0.678	Don't Reject
edu_lvl	39	2.26	0.91	25	1.80	0.82	2.04	62	0.46	0.046	Reject

Note. edu_lvl = highest education level completed. *For this particular table, P-values are specifically reported with three decimal places instead of two, as the values for the age and education level variables are within the third decimal range of the 0.05 decision threshold.

Here, the null hypothesis is that the means across both intervention groups are equal, and the alternative hypothesis is that they are not equal, or in other words, that their difference is not equal to 0. The latter is illustrated by the two-sided P-values of 0.055 and 0.344 for the age and gender means comparisons, and the P-values of 0.678 and 0.046 for the country of origin and education level respectively. While the reported P-value for mean education level comparisons is barely under the 5% statistical significance level and could be rounded at 0.05, all three other P-values are above the 5% decision threshold. The null hypothesis for mean comparisons of age, gender and country of origin cannot be rejected, implying that the age, gender and origin country means across groups do not differ, from a statistically significant point of view. Even though the null hypothesis for means comparison is rejected for the education level variable, the proximity of its P-value to the 5%

threshold as well as the non-significance of mean differences for the other demographic variables allows for a result aggregation across variables. Thus, as the difference between the means of 3 out of 4 demographic variables does not significantly differ across groups with one variable having slightly differing means, it can be stated that overall, the randomization process has functioned.

Concerning treatment group-related descriptive statistics, Table 1.3. elaborates on the total time spent on the survey, the exposure time to the intervention, as well as the manipulation checks, for both Warning and Guidelines group.

Table 1.3. Descriptive statistics of survey length, manipulation check and treatment exposure, by treatment groups

	Treatment W					Treatment G				
Variable	**Obs.**	**Mean**	**S.D.**	**Min**	**Max**	**Obs.**	**Mean**	**S.D.**	**Min**	**Max**
exposure	39	15.91	14.70	6.40	82.29	25	60.59	26.62	22.62	144.46
duration	39	354.33	274.01	121	1404	25	405.08	509.59	158	2688
manip_check	39	7.18	2.70	0	10	25	7.52	1.85	3	10

Note. manip_check = manipulation check, also referred to as the usefulness rating of each treatment.

For the Warning intervention group, the mean exposure to the treatment lasted 15.91s and the mean survey duration was 354.33s or about 5.91 minutes. After the data cleaning process, the shortest exposure accepted is of 6.40s and the longest one of 82.29s. Concerning the Guidelines group, the mean exposure to the treatment lasted 60.59s and the mean survey duration was 405.08s or about 6.75 minutes. Here, the shortest and longest treatment exposures collected are respectively of 22.26s and 144.16s. Considering the difference in length between the Guidelines and Warning interventions, with the former being considerably longer, similar differences in duration and exposure time are to be expected. For the treatment usefulness rating, reported here as the manipulation check, the mean rating is 7.18 for the Warning group and 7.52 for the Guidelines group, displaying a difference of only a few decimal points between the rated usefulness of each treatment. The minimum rating being 0 out of 10 for the Warning group and 3 out of 10 for the Guidelines group can be potentially justified by the presence of more participants in the Warning group, with the higher standard deviation of 2.70 against 1.85 indicating that the ratings are indeed more spread out across the rating scale for the Warning treatment.

Below, Table 1.4. adds some descriptive statistics on news content perception, which is measured at the individual level for each of the news materials. As a reminder, the response value of 1 is assigned to the answer “Not accurate at all”, 2 to “Not very accurate”, 3 to “Somewhat accurate” and 4 to “Very accurate”, when asking the question “To the best of your knowledge, is the claim in the above headline accurate?”.

Table 1.4. Descriptive statistics of news perceptions

Variable	Mean	Stand. deviation	Min	Max
t1	2.78	0.95	1	4
t2	2.09	0.81	1	4
t3	1.70	0.95	1	4
t4	2.60	0.79	1	4
f1	3.06	0.81	1	4
f2	1.63	0.85	1	4
f3	2.00	0.80	1	4
f4	1.73	0.80	1	4

Note. t = true news, f = fake news, with the number after specifying the news headline being referred to.

For t1, t2 and t4, the mean perception rating has the unit value of 2 (with respective values of 2.78, 2.09 and 2.60), implying that the mean evaluation for these true news content is that they are not very accurate. While getting close to response value 2, the mean accuracy rating for t3 is of value 1.70, meaning that the mean accuracy assessment for this is headline is that it is not accurate at all. It is of particular interest here that no true news content reaches the realm of accurate ratings of response values 3 and 4 or in other terms, no true news is perceived as accurate or somewhat accurate.

Regarding fake news headlines, for f2 and f4 the mean accuracy ratings are 1.63 and 1.73 respectively. This implies that both headlines were perceived as being not accurate at all. With a mean accuracy rating of 2.00, f3 was perceived as being not very accurate. Strikingly, f1 has the highest mean accuracy rating of all news headlines, true and false combined. At a mean value of 3.06, f1 was rated as being somewhat accurate.

These values clearly indicate that the news headlines were not perceived by the respondents as intended, as no true news is perceived as being accurate and the individual headline with the highest accuracy rating is a fake news piece, namely f1. Reasons as to why such phenomena could have occurred will be further elaborated upon in the *Discussion* segment.

Planned analysis

To examine the main hypothesis stating that using a forewarning message helps misinformation discernment more than assessing guidelines on how to detect fake news, a Mann–Whitney U-test is performed to compare the effectiveness of the different awareness-raising interventions. The Mann–Whitney test, also referred to as a Wilcoxon rank-sum test, is a non-parametric test used to compare two independent groups with data on an ordinal scale (McCrum-Gardner, 2008). This test compares the medians of the accuracy rating of both groups. As an analysis method, the Mann–Whitney U-test is suitable for this study as fake news discernment is translated to ordinal data measured on a 4-point Likert scale, with the differences between two independent treatment groups being the results that are sought. It is worth noting that while true content accuracy rating is important in context of news discernment and analyzing it can tell a lot about individuals' cognitive judgement abilities, the key variable to answer the research question is misinformation perception. That being mentioned, the analysis lead by both hypotheses focuses on the distinction between true and fake news content, with attention allocated to true news belief adding empirical substance and discussion elements. In other words, the main hypothesis predicts a lower accuracy rating for fake news and a higher rating for true news upon exposure to the Warning intervention, when compared to both ratings in the Guidelines intervention group. Regarding the secondary hypothesis, claiming that older adults show reduced misinformation discernment compared to younger adults, the same analytical method applies because of the nature of the data and the measurement scale remaining the same. The difference to note here is that instead of using the two treatment groups as independent samples, the younger and older age cohorts are employed for this hypothesis. The expected effect for the second hypothesis is that older adults report higher accuracy rating for fake news and lower rating for true news, compared to younger adults.

Another important aspect to discuss is the scale at which the analysis is performed. Instead of aggregating accuracy ratings across fake and true news material, this paper compares the ratings of individual news content, as recommended by the *practical guide to doing behavioral research on fake news and misinformation* by Pennycook et al. This decision can be explained by the overshadowing of variation across individual headlines which tends to occur when averaging ratings between news categories. As a matter of fact, data from Table 1.4. supports this item-level analysis approach when comparing the differences in accuracy ratings for fake news. Ranging from 1.63 for f2 to 3.06 for f1, considerable variances are indeed observable for fake news accuracy assessments.

Results

First hypothesis testing

Table 2.1 and 2.2 below show the outputs of the Mann–Whitney U-test for the primary hypothesis, for true news content and fake news content respectively. The null hypothesis is that the accuracy rating of the Warning treatment group is equal to the accuracy rating of the Guidelines treatment group and the alternative hypothesis is that there is a difference between the former and the latter. Concerning the probability reported in the final row, it is obtained by pairing each individual news accuracy rating from the Warning group with those from the Guidelines group, counting the number of times that the accuracy rating from the Warning group is greater than the rating from the Guidelines group, then dividing the result by the total number of pairs.

Table 2.1. Mann-Whitney test statistics for true news between treatment groups

	t1	t2	t3	t4
Rank sum (group_W)	1141.50	1095.50	1192.50	1284.50
Rank sum (group_G)	938.50	984.50	887.50	795.50
z	-1.82	-2.53	-1.16	0.26
P-value	0.07	0.01	0.25	0.80
P(t(group_W)> t(group_G))	0.37	0.32	0.42	0.52

Note. t = true news, with the number after specifying the news headline being referred to. Variables *group_W* and *group_G* refer to the Warning treatment group and the Guidelines treatment group respectively.

The reported P-values are respectively 0.07 for t1, 0.01 for t2, 0.25 for t3 and 0.80 for t4. While t1 gets very close to reaching a P-value below the 0.05 decision level, only t2 out of the 4 headlines reaches a P-value under that threshold. This means that ¾ of these results are statistically non-significant, at a confidence level of 95%, thus rejecting the null hypothesis stating that true news accuracy ratings for both treatment groups are equal. In other terms, no conclusive effect from the treatments on true news accuracy rating can be observed.

While taking that into account, it is still worth examining the last row, depicting the probability that the truthfulness assessment of the material presented is higher for treatment group W than for treatment group G. As this probability is under 0.50 for the first 3 news materials, this entails that on average, group G perceived true news more accurately than group W for 3 out of 4 headlines presented. Specifically, true news were more accurately perceived by group G than W with probabilities of 0.63 for t1, 0.68 for t2 and 0.58 for t3. Regarding the 4th news content, this probability is only slightly higher than 0.50 in favor of group W, thus not changing by much the overall support for group G's discernment ability. Drawing from the last row only, group G appears to have a better true information accuracy rating than group W.

Table 2.2. Mann-Whitney test statistics for fake news between treatment groups

	f1	f2	f3	f4
Rank sum (group_W)	1209	1341.50	1097	1203.50
Rank sum (group_G)	871	738.50	983	876.50
z	-0.86	1.15	-2.50	-0.95
P-value	0.39	0.25	0.01	0.34
P(f(group_G)> f(group_W))	0.44	0.58	0.33	0.43

Note. t = true news, with the number after specifying the news headline being referred to. Variables *group_W* and *group_G* refer to the Warning treatment group and the Guidelines treatment group respectively.

Regarding fake news assessment, the P-values attached to each news content are 0.39 for f1, 0.25 for f2, 0.01 for f3 and 0.34 for f4. When examining the statistical significance of these values, it is observed that only f3 out of the 4 headlines is statistically significant at a confidence level of 95% (P-value < 0.05). Thus, here as well, the null hypothesis stating that accuracy ratings for both treatment groups are equal is rejected and no conclusive effect from the treatments on fake news accuracy rating can be observed.

The last row of Table 2.2 implies that for 3 out of the 4 misinformation material presented to both groups, group G appraises the material more accurately than group W with probabilities of 0.56 for f1, 0.67 for f3 and 0.66 for f4. As for the true news rating from Table 2.1, only one news material appraisal favors group W, namely f2, leaving the overarching support in favor of group G's fake news perception abilities.

When comparing these results with the prediction made in the first hypothesis, namely that using a forewarning message helps misinformation discernment more than assessing guidelines on how to detect fake news, this hypothesis is rejected based on the lack of statistically significant results for both fake and true news. Thus, the expected effect of awareness-raising interventions on misinformation discernment cannot be scientifically backed up by these results.

Second hypothesis testing

Table 2.3 and Table 2.4 below display the empirical data and analyses conducted to test the secondary hypothesis. The null hypothesis states that the accuracy assessment of the younger age group is equal to the accuracy assessment of the older age group and the alternative hypothesis is that there is a difference between both groups' accuracy assessment. For the probability reported in the last row, it is obtained by pairing each individual news accuracy rating from the younger age group with those from the older age group, counting the number of times that the accuracy rating from the younger age group is greater than the rating from the older age group, then dividing the result by the total number of pairs.

Table 2.3. Mann-Whitney test statistics for true news between age groups

	t1	**t2**	**t3**	**t4**
Rank sum (younger)	1161	1321	1249	1065
Rank sum (older)	919	759	831	1015
z	1.71	4.03	3.15	0.37
P-value	0.09	0.00	0.01	0.71
P(t(younger)> t(older))	0.62	0.77	0.70	0.52

Note. t = true news, with the number after specifying the news headline being referred to. Variables *younger* and *older* refer to the younger and older age groups respectively.

Inspecting the P-values of these results, it is observed that they are significant for t2 and t3 at a confidence level of 95% as their P-values are of 0.00 and 0.01. While data for t4 is statistically non-significant due to its P-value of 0.71, data for t1 has a P-value of 0.09, coming very close to the 0.05 threshold for statistical significance. Based on such proximity to a statistically significant P-value level, the null hypothesis claiming that for true news, the accuracy rating of the younger age group is equal to the accuracy rating of the older age group is not rejected. This allows the interpretation of the following results as overall statistically significant, while keeping some skepticism towards their actual accuracy.

The final row illustrates the probability that the truthfulness assessment of the content is higher for the younger age group than for the older age group. This is the case for all of the true news headlines presented, with probabilities of 0.62 for t1, 0.77 for t2, 0.70 for t3 and 0.52 for t4. Therefore, it appears that for true news, the younger age cohort assesses information more correctly than the older age cohort.

Table 2.4. Mann-Whitney test statistics for fake news between age groups

	f1	f2	f3	f4
Rank sum (younger)	958	1027.50	1123	1128
Rank sum (older)	1122	1052.50	957	952
z	-1.17	-0.19	1.19	1.28
P-value	0.24	0.85	0.23	0.20
P(f(younger)> f(older))	0.42	0.49	0.58	0.59

Note. f = fake news, with the number after specifying the news headline being referred to. Variables *younger* and *older* refer to the younger and older age groups respectively.

Fake news assessments depicted in Table 2.4 is very nuanced between different age groups. For f1 and f2, the probability that the older group assesses misinformation more accurately is higher than for the younger group, at 0.58 and 0.51 respectively. This effect is reversed for f3 and f4, where the younger group has a higher probability than the older one of accurately assessing the presented fake news, with probabilities of 0.58 and 0.59 respectively. These mixed effects lose some of their importance when put in context with the statistical significance of the observed results, as for all 4 of the presented headlines, the P-values are above the threshold of 0.05. The null hypothesis stating that for fake news, the accuracy assessment of the younger age group is equal to the accuracy assessment of the older age group is rejected. Thus, on the basis of these statistically non-significant results, no conclusive effect on how younger adults assess misinformation compared to older adults can be drawn.

While it is observed that younger adults display better news discernment than older adults in regard to true news content, such observation cannot be made regarding the fake news material. Therefore, the secondary hypothesis stating that older adults show reduced misinformation discernment compared to younger adults, is rejected based on the lack of statistically significant results for fake news.

Analysis with outliers

In order to test the validity of the results found for both hypotheses, the Mann-Whitney tests for both treatment groups and age cohorts are performed a second time, this time including the 11 outliers eliminated through the data cleaning process by the time criteria and that could not be ensured a satisfying exposure to the treatments. Tables A1 to A4 report these results in the Appendix section of this paper.

Comparing Tables 2.1 and 2.2 with A1 and A2, it is observed that by including the outliers, the few statistically significant news headlines assessments, namely t2 and f3, become statistically non-significant at a 95% confidence level, by both moving from P-values of 0.01 to 0.09 and 0.28 respectively. This has the effect of removing all nuance and rendering the entirety of the results presented in both Tables 3.1 and 3.2 statistically non-significant. Thus, for the first hypothesis, removing outliers has a constructive impact on the data as it allows results to be reported more accurately.

Contrasting Tables 2.3 and 2.4, with A3 and A4, the P-values of two individual news pieces changes from being statistically non-significant to statistically significant at a 95% confidence level. This is the case for t1, whose P-value decreases from 0.09 to 0.02 and for f4, where the value declines from 0.20 to 0.04. Along with these changes, the probability of the younger age cohort having more accurate news content ratings than the older age group increases for the same two headlines. Regarding the secondary hypothesis, while including the outliers slightly increases the scientific support of younger adults having more accurate news perception than older adults, this is not enough to accept the hypothesis, as observations for other individual news content appraisals cannot be scientifically backed up collectively. Therefore, outliers removal is justified and including them does not add empirical relevance to the results.

Discussion

Noting the lack of statistical significance of the first hypothesis' results, hindsight in the methods employed and the news material itself is required. Looking at the size of the effect, observations on the probability of one intervention being more favored than the other are mixed across individual headlines. Put in other words, no statistically significant support in favor of one specific intervention can be pointed out. While no other paper compares these specific interventions, academic literature has previously commented on the idea of dividing inoculation theory's core elements, the forewarning and the refutational preemption process. Pfau (1992) mentions the necessity of both aspects applied in tandem for the proper functioning of inoculation, insisting on each element having a pivotal role. Therefore, the initial idea of comparing these two specific interventions might be at the root of the lack of empirically significant results for the effect of each treatment. Recognizing this also puts in perspective the specific advocacy in favor of the effects of employing a forewarning intervention alone, as formulated in the first hypothesis. With the main arguments backing the forewarning's potential being the lengthy exposure and active information processing requirements of inoculation theory, the latter is after all, in its entirety, a more consistent and scientifically motivated method to oppose misinformation than a separate forewarning message.

On a more positive note, when looking solely at its direction, the effect of awareness-raising interventions on misinformation perception is more inclined on supporting the impact of the Guidelines treatment for both fake and true news ratings. Such observations resonate with the previous work from Lutzke et al. (2019), in which after being exposed to a similar guideline ranking intervention, subjects were found to consistently less share and trust Facebook headlines about climate change misinformation. Yet another study backing up the inclusion of misinformation detection tips as a fake news fighting method is the study conducted by Guess et al. (2020). In this paper, media literacy tips were shown to be effective at diminishing the perceived accuracy of political fake news at an international scale. Thus, while this paper's data can only hint at and not draw an empirically robust conclusion, the inclusion of media literacy guidelines appears to be promising in combatting misinformation.

Concerning the second hypothesis, results for true news content favor the truthfulness rating of the younger age group. Such observation follows and builds on remarks made in studies such as that of Hobbs & Jensen (2009), indicating that there is a gap between how younger people and older adults interact with and perceive digital media and noting the need for future research to explore the relationship between media literacy and the younger generations. While the focus of this paper is not to explain said relationship, data reported in Table 2.3. provides some answer elements in the context of COVID-19 information judgement. Regarding the results for fake news from Table 2.4., no conclusive effect of inclusion in a specific age cohort on accuracy ratings can be observed. A first possible cause of such statistically non-significant observations resides in the presence of an exogenous variable, not accounted for in this analysis, which impacts headlines' accuracy assessment. Another explanation along these lines, applicable to both the first and second hypotheses is given below.

To explain the statistically non-conclusive results derived for both hypotheses, a closer look should be taken at the news perception statistics reported in Table 1.5. As every single true news headline was interpreted as fake news and one fake news headline, f3, showed the highest accuracy rating of all news combined, it can be inferred that some biases were involved in the content selection process. Such biases may have come from the focus on including relevant yet somewhat specific sources with medical-related content. Deriving from that, further analysis involving the impact of the highest education level completed on accuracy ratings could have been led, with particular interest given to the specific headlines containing specific medical elements. Yet another potential reasoning contributing to the observed statistically non-conclusive results resides in the 4-point Likert scale format used in the survey and inspired by the research guide of Pennycook et al. (2021). By design, 4-point Likert scales force participants to express an opinion. The purposefully excluded "I don't know" answer point impacts subjects' judgements in the sense that they are restricted to being opinioned about each single headline, even when they might not understand several key elements of the headline, or the message portrayed. For this reason, misreports of accuracy ratings may have occurred at several instances.

Finally, for the second hypothesis, the assumption that the experimental manipulation does not impact the accuracy ratings of both age groups cannot be totally controlled for in this paper. Ensuring that would require the presence of a control group besides the 2 treatment groups, and a comparison of accuracy ratings of younger adults against older adults, between treated and non-treated subjects. Including a third group, controlling for treatment exposure, would strongly improve the internal validity of this study's results, but also requires more in-depth statistical analyses as well as a significantly larger number of participants. This would affect the feasibility of this work at the benefit of more accurate measures.

Conclusion

The scope of this paper was to empirically assess and compare the impact that different awareness-raising methods can have on misinformation perception, with specific attention to COVID-related fake news. Contrasting an inoculation-based media literacy tips intervention with a short and concise warning message, discernment between true and fake COVID-19 news headlines was empirically analyzed through the exposure to one of the two informative treatments. The first hypothesis stated that using a forewarning message helps misinformation discernment more than assessing guidelines on how to detect fake news. With the help of a Mann-Whitney test, data collected on news accuracy rating was compared across 8 news headlines individually, and as most of the results lacked statistical significance, the first hypothesis was rejected. For the second hypothesis claiming that older adults show reduced misinformation discernment compared to younger adults, a similar Mann-Whitney test was performed at an individual news piece level. With stronger true news distinguishment being observed for the younger group, this effect was not present in regard to fake news content. The second hypothesis was therefore rejected due to the expected effect not being present for both true and fake news content.

The research question: "How can exposure to different awareness-raising methods influence COVID-19 fake news perception?" is addressed as follows: no reliable effects were found for the two awareness-raising techniques tested to combat fake news perception in this paper. In other terms, the interpretation of the scientific methods employed does not indicate statistically significant support in favor of either the use of a forewarning message or media literacy tips. As this conclusion is influenced by the previously discussed limitations of this research, various recommendations for future research on this topic can be provided.

The main theoretical implication of this study is for future literature employing inoculation theory to not contrast its two core elements, but rather to inoculate as defined in procedure: by using a forewarning message accompanied by a refutational preemption. If for whatever reason, future researchers seek to contrast these two elements within a different setting, it is suggested to have a larger sample and at least one control group, to compare the interventions' efficacy against the status quo, no other treatment. Following on the necessity of a control group, its presence would also ensure higher internal validity of the obtained results, making the observed effects more reliable. As age was shown to have a significant effect on COVID-19 true news material perception, further studies could test the validity and reliability of this effect in different settings. For instance, education-level and origin country variables could be included in the analysis of this effect, to check whether more educated young adults or young adults born in specific areas still assess COVID-19 true information more accurately than their older relatives. This effect of age group on true news accuracy perception can also be tested for other relevant news categories (such as climate change information) and by employing different methods, some of which are suggested below.

Adding on to practical implications, employing different headlines material, focusing on other aspects than medical relevancy can be of great pertinence. For example, future papers using COVID-19 headlines could use a greater number of specific vaccines-related articles, to explore whether vaccines issued by distinct companies are associated with different trust levels. To minimize biases during the content selection process, performing a baseline test on the selected news material could be of help for following studies. This would consist of measuring pre-treatment accuracy perception-related characteristics of the participants, with the benefit of adding a temporal factor to the research, where before and after treatment measures can be contrasted, in the manner of clinical trials. Pursuing the methodology recommendations, employing different measures of accuracy ratings can also be pertinent for future papers. This can include utilizing Likert scales with an uneven number of points, such as 5-point or 7-point scales, allowing respondents to express their lack of knowledge on a particular topic.

Bibliography

Adair, J. G. (1984). The Hawthorne effect: a reconsideration of the methodological artifact. *Journal of applied psychology*, 69(2), 334.

Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. *Communication Monographs*, *77*(3), 281-311.

Bertram, D. (2007). Likert Scales… are the meaning of life: *CPSC 681–Topic Report,* 2(10), 1-10.

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current directions in psychological science*, 29(3), 316-323.

Compton, J. (2013). Inoculation theory. *The SAGE handbook of persuasion: Developments in theory and practice*, 2, 220-237.

Compton, J., & Pfau, M. (2004). Use of inoculation to foster resistance to credit card marketing targeting college students. *Journal of Applied Communication Research*, 32(4), 343-364.

Cook, J., Lewandowsky, S., & Ecker, U. K. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PloS one*, *12*(5), e0175799.

Flash Eurobarometer 464: Fake News and Disinformation Online (v1.00). (2018). [Data set]. *European Commission, Directorate-General for Communication*. Retrieved from http://data.europa.eu/88u/dataset/S2183_464_ENG

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances*, 5(1).

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (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 of the United States of America*, 117(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Hobbs, R., & Jensen, A. (2009). The past, present, and future of media literacy education. *Journal of media literacy education,* 1(1), 1.

Islam, M. S., Kamal, A. H. M., Kabir, A., Southern, D. L., Khan, S. H., Hasan, S. M., ... & Seale, H. (2021). COVID-19 vaccine rumors and conspiracy theories: The need for cognitive inoculation against misinformation to improve vaccine adherence. *PloS one*, 16(5), e0251605.

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, 58, 101964.

McCrum-Gardner, E. (2008). Which is the correct statistical test to use?. *British Journal of Oral and Maxillofacial Surgery*, 46(1), 38-41.

Mukhtar, K., Javed, K., Arooj, M., & Sethi, A. (2020). Advantages, Limitations and Recommendations for online learning during COVID-19 pandemic era. *Pakistan journal of medical sciences*, 36(COVID19-S4), S27.

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology*, 7(1), 25293.

Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020a). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11), 4944-4957.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020b). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science*, 31(7), 770-780.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521-2526.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388-402.

Pfau, M. (1992). The potential of inoculation in promoting resistance to the effectiveness of comparative advertising messages. *Communication Quarterly*, 40(1), 26-44.

Pfau, M., Bockern, S. V., & Kang, J. G. (1992). Use of inoculation to promote resistance to smoking initiation among adolescents. *Communications Monographs*, 59(3), 213-230.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Van Der Linden, S. et al. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society open science*, 7(10), 201199.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), 99-118.

Snopes. (2022). *Red Flags: How to Identify Suspicious Rumors*. Retrieved April 10, 2022, from https://www.snopes.com/articles/401830/red-flags-suspicious-rumors/

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008.

World Health Organization. (2020). WHO COVID-19 case definition (No. WHO/2019-nCoV/Surveillance_Case_Definition/2020.1). *World Health Organization*.

Wylie, L. E., Patihis, L., & McCuller, L. L. (2014). Misinformation effect in older versus younger adults: A meta-analysis and review. *The elderly eyewitness in court*, 52-80.

Appendix

Figure A1. First fake news headline

Figure A2. Second fake news headline

Figure A3. Third fake news headline

Figure A4. Fourth fake news headline

Figure A5. First true news headline

Figure A6. Second true news headline

Figure A7. Third true news headline

Figure A8. Fourth true news headline

Table A1. Mann-Whitney test statistics for true news between treatment groups (with outliers)

	t1	**t2**	**t3**	**t4**
Rank sum (group_W)	1689.50	1643	1770	1899.50
Rank sum (group_G)	1160.50	1207	1080	950.50
z	-1.10	-1.66	-0.19	1.35
P-value	0.29	0.09	0.85	0.18
P(t(group_W)> t(group_G))	0.43	0.39	0.49	0.59

Note. t = true news, with the number after specifying the news headline being referred to. Variables *group_W* and *group_G* refer to the Warning treatment group and the Guidelines treatment group respectively.

Table A2. Mann-Whitney test statistics for fake news between treatment groups (with outliers)

	f1	**f2**	**f3**	**f4**
Rank sum (group_W)	1734	1917.50	1692	1782
Rank sum (group_G)	1116	932.50	1158	1068
z	-0.60	1.63	-1.09	-0.05
P-value	0.55	0.10	0.28	0.96
P(f(group_G)> f(group_W))	0.46	0.60	0.43	0.50

Note. t = true news, with the number after specifying the news headline being referred to. Variables *group_W* and *group_G* refer to the Warning treatment group and the Guidelines treatment group respectively.

Table A3. Mann-Whitney test statistics for true news between age groups (with outliers)

	t1	t2	t3	t4
Rank sum (younger)	1685.50	1845.50	1827.50	1548
Rank sum (older)	1164.50	1004.50	1022.50	1302
z	2.25	4.09	4.02	0.76
P-value	0.02	0.00	0.00	0.45
P(t(younger)> t(older))	0.65	0.76	0.75	0.55

Note. t = true news, with the number after specifying the news headline being referred to. Variables *younger* and *older* refer to the younger and older age groups respectively.

Table A4. Mann-Whitney test statistics for fake news between age groups (with outliers)

	f1	f2	f3	f4
Rank sum (younger)	1398	1501	1631.50	1661
Rank sum (older)	1452	1349	1218.50	1189
z	-0.94	0.23	1.68	2.04
P-value	0.35	0.82	0.09	0.04
P(f(younger)> f(older))	0.44	0.51	0.61	0.63

Note. f = fake news, with the number after specifying the news headline being referred to. Variables *younger* and *older* refer to the younger and older age groups respectively.