Are Media Professionals Better at Fake News Recognition and Less Susceptible to Confirmation Bias? A Signal-Detection Approach

Research Report

Laboratory for Social & Cognitive Informatics

2022

Abstract

Belief in disinformation can be driven by various factors. Messages are perceived as more credible if they support personal beliefs—this effect has been known as confirmation bias. On the other hand, media literacy is likely to be an effective measure against users' susceptibility to fake news, and, supposedly, to individual biases. However, little is known about how disinformation is perceived by expectedly the most competent audience—media professionals. By conducting an online experiment (N=1946) in Russia in 2021, we test how confirmation bias affects trust in fake and true news about three socially divisive topics (LGBT, abortions, and death penalty) among two groups of participants: media professionals and ordinary social media users. Our study shows a strong effect of confirmation bias on the perceived news credibility across all topics regardless of media professionalism. We also find that media professionals are indeed better at fake news detection than other users, which, however, is mostly explained by their advanced fact-checking skills. If the participants did not verify news during the experiment, the news discrimination was almost the same in both groups, but if they did, media professionals discriminated news articles much better than ordinary users. Finally, we find that the news that has been reported as seen prior to the experiment is perceived as more credible. Thus, our study shows that while confirmation bias and familiarity significantly affect trust in the news (irrespective of its veracity) even among media professionals, high-quality fact-checking can reduce susceptibility to fake news. This suggests that further dis- and misinformation research, rather than testing the effects on participants' abilities to guess fakes in isolated experimental environments, should proceed to investigate the efficiency of fact-checking strategies available in real-world online settings.

Introduction

In a so-called post-truth era, both disinformation (deliberate spread of falsehoods) and misinformation (unintentional spread of misleading information) can have a significant impact on society. The term “fake news”— here defined as fabricated news articles created with the intention to deceive (Allcott and Gentzkow, 2017; Lazer et al, 2018)—has been widely used by researchers, journalists, the public, and politicians over the last years. Although some researchers call the fake news discourse “an alarmist narrative” (Altay, Berriche & Acerbi, 2021) suggesting that this problem is overrated, mis- and disinformation can nevertheless cause harm to society in various ways. By internalizing inaccurate information individuals might make suboptimal decisions based on the wrong evidence (Rapp & Salovich, 2018). This, in turn, can lead to changes in voter behavior (Zimmermann & Kohring, 2020) or risky health decisions such as refusal to get vaccinated (Carrieri, Madio & Principe, 2019; Loomba et al, 2021).

The spread of disinformation has become a concern both for the public (Newman et al, 2020) and professional journalists whose routine has been affected by the necessity to verify false claims and debunk online rumors (PEN America, 2022). As a response to this concern, numerous studies have started exploring factors affecting the perceived credibility of fake news including individual differences and message characteristics (Bryanov & Vziatysheva, 2021). Yet, while the audience’s susceptibility to fake news has been widely studied, the way such messages are evaluated by professional journalists and editors has not been a focus of any fake news research.

This study explores how confirmation bias interplays with media literacy in the context of news evaluation. We test the effect of confirmation bias on two types of audiences—ordinary social media users and media professionals—by presenting them fake and true news articles about socially controversial topics: abortion, LGBT rights, and the death penalty. To our knowledge, this is the first online experiment that examines how media professionals perceive disinformation.

Methodologically, credibility ratings of true and fake news confound respondents’ discriminability about the news topic and respondents’ general response tendencies to accept or reject a news item as “true”. Discriminability is affected by respondents’ knowledge and the degree of similarity between fake and true news. To clearly separate news discriminability from respondents’ response biases, we base our analysis on the signal detection theory (SDT) adopted from psychological memory research (Green & Swets, 1966).

Confirmation bias in news consumption

Confirmation bias is commonly understood as a tendency to seek or interpret information in ways that support “existing beliefs, expectations, or a hypothesis in hand” (Nickerson, 1998). The idea of this process originates in Festinger's concept of cognitive dissonance (1957)—a state of psychological discomfort caused by inconsistencies between the existing knowledge and new information. Confirmation bias motivates individuals to omit messages that challenge their attitudes (Knobloch-Westerwick et al, 2015) and, thus, helps them to avoid cognitive dissonance. Nickerson (1988) notes that this is not an explicit or deliberate process but rather “an unwitting selectivity” of evidence that proves people’s beliefs.

Confirmation bias may affect individual decisions or information evaluation in a variety of contexts—from interpersonal communication to policy making. Taber and Lodge (2006), who asked respondents to assess arguments concerning affirmative action and gun control, found that people tend to more actively seek arguments supporting their opinion than those opposing it. Scholars also provide evidence that while individuals easily accept attitudinally congruent arguments, they eagerly counterargue and denigrate the statements that do not match their opinion (disconfirmation bias). Čavojová, Šrol, and Adamus (2018), who focus on a more narrow type of confirmation bias—myside bias,—came to similar results using the syllogisms evaluation task: they found that participants had difficulties accepting logically valid conclusions that contradicted their attitudes and rejecting logically invalid conclusion that confirmed their attitudes.

The same reasoning processes are activated when individuals consume the news. Confirmation bias can influence news selection and interpretation. For instance, experiments examined the role of confirmation bias in selective exposure to information (e.g., Westerwick, Johnson & Knobloch-Westerwick, 2017; Knobloch-Westerwick, Johnson & Westerwick, 2015; Knobloch-Westerwick et al, 2015; Knobloch-Westerwick & Meng, 2009). Participants of these studies received a set of messages—for example, in the form of a newsfeed or an online magazine—and were asked to browse the articles. Overall, these studies show that people prefer and spend more time with attitude-consistent than with attitude-discrepant information.

Confirmation bias may affect not only the choice of information to consume but also the likelihood to believe it. Studies demonstrate that news is perceived as more trustworthy when it aligns with pre-existing beliefs. Kim and Dennis (2019) found that users were more likely to read, like, and share articles that support their political views (left- or right-leaning in the case of this study). In the context of fake news, this bias may increase the susceptibility to false information that supports people's beliefs and, conversely, lower trust in true news that presents a conflicting point of view. In another experiment, Moravec, Minas and Dennis (2018) used EEG to examine people's reaction to fake news and found that participants pay more attention to attitude-consistent headlines while ignoring attitude-discrepant ones. Furthermore, the authors observed a neurophysiological indication of cognitive dissonance: if attitude-congruent headlines were labeled as fake, participants engaged in extra cognitive activity. However, this additional cognition still did not force them to reject such news as false.

In line with the existing research, our first hypothesis suggests that:

H1: Irrespective of their truth status, news articles aligned with one's attitudes will be perceived as more credible than attitudinally discrepant articles.

Media literacy

Scholars and experts have been actively emphasizing the importance of media literacy skills for news consumers. In 2020, the Council of the European Union adopted "The Council conclusions on media literacy in an ever-changing world". The document acknowledges the necessity to equip citizens with media literacy and critical thinking because of the growing exposure to disinformation "especially in times of major global crisis, such as the COVID-19 pandemic" (Council of the European Union, 2020). One of the common definitions of media literacy is ''the ability to access, analyze, evaluate and communicate messages in a variety of forms'' (Aufderheide, 1993). European Commission (2007) identifies several levels of media literacy, one of which is “having a critical approach to media as regards both quality and accuracy of content”.

Existing research, indeed, provides promising evidence that media literacy interventions have a positive impact on fake news recognition. For example, a cross-national experiment conducted in the USA and India shows that media literacy interventions improve users’ ability to discern between true and false news headlines (Guess et al, 2020). Amazeen and Bucy (2019) demonstrate that a better understanding of professional news operations and procedures is associated with lower susceptibility to fake news. Likewise, another study, by Moore and Hancock (2022), reveals that a digital media literacy intervention (i.e., an interactive course) helps older adults to better recognize false information. Yet, some studies show different results: thus, Jones-Jang, Mortensen, and Liu (2019) find that neither media literacy, nor news literacy is a significant predictor of better fake news discernment, although information literacy is.

However, to our knowledge, no research so far has experimentally tested how professional media competence (e.g., working in journalism) is related to the ability to distinguish between fake and true news. This is surprising given that news media plays a significant role in the misinformation problem: first, journalists can unintentionally share inaccurate or false facts due to failed verification (e.g., Silverman, 2015); second, by debunking fake news, media outlets can raise public awareness of such stories (Tsfati et al, 2020).

Verification and evaluation of the credibility of information are one of the key journalistic practices. Of course, in real-life media production, not all information is getting properly fact-checked, which can happen because of time constraints (Himma-Kadakas, 2017), lack of verification skills (Brandtzaeg et al, 2016), desire to generate traffic and shares (Silverman, 2015), or simple ignorance (Saldaña & Vu, 2021). Still, we assume that media professionalism will be a significant predictor of the ability to recognize fake news, which leads us to the following hypothesis:

H2: Media professionals are more accurate at discriminating between fake and true news than ordinary social media users.

From the normative perspective, journalists should adhere to the value of objectivity, thus, being impartial, objective, and credible (Deuze, 2005). Thus, we expect that impartiality will ultimately lead to a lower level of confirmation bias among media professionals, which makes us assume the following:

H3: Media professionals are less susceptible to confirmation bias than ordinary social media users.

User comments

News on social media are often forwarded along with evaluative comments expressing attitude towards the topic. Drawing on the idea that social media facilitate the formation of echo chambers (Quattrociocchi, Scala & Sunstein, 2016), we assume that such comments may amplify confirmation bias when the comment aligns with the valence of the news or reduce its effect when the comment is discrepant with the news valence. A similar argument can be made for an alignment between comment and participants’ attitude towards the issue. Thus, the following hypotheses are proposed:

H4a: Confirmation bias effect will be stronger if the news is aligned with the reader's attitude and is accompanied by a comment expressing the similar attitude.

H4b: Confirmation bias effect will be weaker if the news is aligned with the reader's attitude but is accompanied by a comment expressing the conflicting attitude.

Signal detection theory

Porshnev, Rabe, Terpilovskii, and Kliegl (2022; see also Batailler et al, 2022) use Signal Detection Theory (SDT; Green & Swets, 1966) for the simultaneous estimation of experimental effects on discriminability of true and fake news and response bias from credibility ratings. Here it is important to introduce a distinction between response bias and confirmation bias. Response bias is any systematic shift in participants' answers that could be associated with different variables (e.g., age, news veracity, etc.), whereas confirmation bias is a specific case of response bias occurring when information is aligned with a person's attitude.

SDT is a well-established theoretical framework in experimental psychology that was initially designed to accurately measure humans' ability to discriminate auditory pure-noise stimuli from stimuli in which a weak tone signal was presented along with the same noise (e.g., Tanner & Swets, 1954). Later SDT was implemented in a variety of contexts—from studying weather forecasting behavior (Harvey et al, 1992) to deception detection among law enforcement investigators (Meissner & Kassin, 2002). Wixted (2020) provides a recent account of the history of SDT. Most relevant for present purposes, the SDT approach can be generalized to memory experiments (Green & Swets, 1966; Taub, 1965) where the task is to discriminate between items not seen or learned before (i.e., new items or "noise") and items that had been seen or learned before (i.e., old items or "noise+signal"). In analogy to this memory-research paradigm, fake news can be considered as new items ("noise") and true news as old items ("noise+signal"). The unique contribution of SDT is that conventional credibility ratings of true and fake news can be used to dissociate readers' accuracy in news discriminability (e.g., knowledge due to professional expertise, the similarity of fake and true news) from response bias to accept or reject an item as true news with confirmation bias being a special case.

Method

Experimental design

The design of this web experiment comprised an orthogonal manipulation of the news topic (3: abortion vs. LGBT vs. death penalty) x veracity of news (2: fake vs. true) x news valence (2: positive vs. negative). Each participant received 12 unique news items (randomly selected from a set of 24 items) that contained an equal number of true and fake news with positive and negative valence on each topic. In addition, each item was accompanied by a negative, neutral, or positive user comment counterbalanced with item valence.

Recruitment

Participants were recruited from among Russian Internet users in three steps in the period between June and July 2021. First, we used the Facebook ads manager system (at the time of data collection, Meta was still legal in Russia) to recruit media professionals on Facebook and Instagram. For that, we designed specific ads and adjusted the settings to target them at people sharing relevant interests (e.g., "journalists", "journalism", etc.). However, since the

Facebook settings (especially for the Russian segment) did not allow customizing ads with great precision based on profession, this step did not bring us many relevant participants. In the second stage, we recruited media professionals using a snowball approach by posting information about the experiment in Facebook groups, Telegram channels, and professional media outlets (e.g., “Zhurnalist” magazine) for Russian media specialists.

Finally, in the last stage, we collected data from the average social media users. We also used the Facebook ads manager system, however at this point we did not include any special interests in the ad settings. While targeting users on Facebook and Instagram, several versions of ads were used; all of them included short texts and a picture. To control the number of respondents based on their socio-demographic characteristics (gender, age, education, and region), we applied a monitoring system that allowed us to adjust the ads settings if some groups were overrepresented.

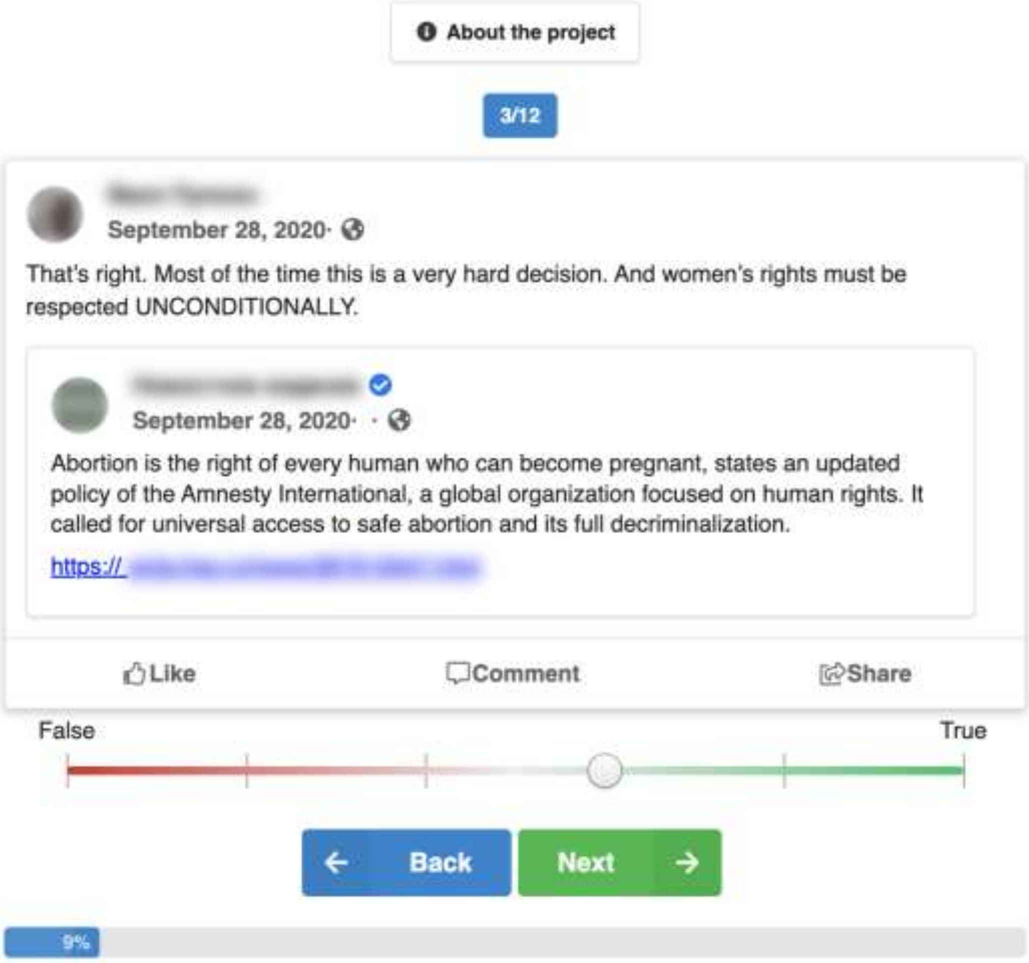
Participants

In total, 1,946 participants (902 female, 948 male, 96 did not indicate their gender) ranging between 21 and 65 years of age (M: 37.8, SD: 10.2) completed the study. Of them, 491 participants worked in the media at the time of the experiment or had worked in the media in the past; 1455 users had no professional media experience.

Since—as expected—the subsample of media professionals was skewed in terms of the level of education (mostly higher), age (mean = 36.2, SD = 10.2), and region (mostly from the two largest cities: Moscow and Saint Petersburg), the subsample of ordinary users was recruited with the aim of achieving a similar demographic pattern. However, the design of this study did not allow us to recruit perfectly matched subsamples. As media professionals were also recruited via the ads for average users in the third recruitment stage, each new respondent in this small group again skewed the distribution.

Stimuli

Participants were exposed to the news items simulating a social media post shared by an unknown user (see Figure 1 for a screenshot).



[redacted]
September 28, 2020·

That's right. Most of the time this is a very hard decision. And women's rights must be respected UNCONDITIONALLY.

[redacted]
September 28, 2020· ·

Abortion is the right of every human who can become pregnant, states an updated policy of the Amnesty International, a global organization focused on human rights. It called for universal access to safe abortion and its full decriminalization.

https:// [redacted]

Figure 1. Screenshot of an item as presented (translated from Russian).

The overall set of stimuli material included 24 fake and true news items on three socially divisive topics: LGBT, abortions, and the death penalty (find examples in the *Appendix*). The attitudes towards these topics can be indicative of a particular set of values that an individual shares. For example, the World Values Survey uses attitudes towards abortions and LGBT as characteristics of the dimensions of global cultural variation. In this framework, a negative/positive attitude towards abortions is attributed to people holding traditional vs. secular-rational values, whereas high/low tolerance to LGBT people indicates the self-expression vs. survival dimension of values (Inglehart, 2018).

True news items were taken from the news outlets and double-checked in other sources. Fake news items were created by a member of the research team with professional training in journalism. The decision to construct fake news was made primarily due to the lack of relevant fake news articles meeting the necessary criteria (topic, valence, etc.) in the actual media environment (Vziatysheva et al, 2021). Furthermore, it allowed us to avoid “an illusory truth effect” in case participants had seen fake news before and to mitigate the confounding effects of stylistic differences that may exist between true and “real” fake news. For the latter goal, fake news articles have been written in a stylistic manner similar to true news. All fake news items and part of the true news items were tested on seven journalists (including editors, reporters, and an editor-in-chief) from five Russian-language media outlets (including digital newspapers, TV, and radio) who were involved in the news production as part of their everyday routine. With this manipulation check, we were able to reject the assumption that professional journalists would recognize fake news without hesitation based on their knowledge of the agenda.

All news items varied by their valence. Building on the Knobloch-Westerwick et al. (2015) methodology, articles were divided into “pro-topic” (e.g., legalization of same-sex marriage) or “contra-topic” (e.g., prosecution of LGBT people).

Since the news genre to a large extent employs neutral stylistics, some of the news articles were accompanied by a user comment which was supposed to amplify (or mitigate) the valence of the news. A unique set of two (positive and negative) comments corresponded to each news item. All user comments were taken from real social media posts either describing the same event (for true news) or related to a similar topic (for fake news).

Procedure

The instrument was adopted from the cross-national study of the effect of media narrative on the perceived news credibility (Bryanov et al., 2022) and adjusted for the new hypotheses and stimuli. Participants recruited via Facebook accessed the experimental interface on the stand-alone website. The first page included a brief description of the task and links to more detailed information about the study. After participants pressed the “start” button, the experiment proceeded as follows:

News evaluation. Participants were shown 12 news items subject to the constraint described under *Experimental design*. The order of the news items was randomized for each user to avoid possible sequence effects. Participants answered whether they considered the information provided to be true or false including also a confidence rating of their judgment.

Questionnaire. After the news evaluation task, participants answered a questionnaire regarding their attitude to the topics, familiarity with items, and whether they checked the veracity of items before providing a rating. They also answered questions about their socio-demographic characteristics, political views, and habits of news consumption. Respondents received different versions of the questionnaire based on their involvement in media production: former and current media employees were given additional questions about their professional experience.

Measures

Perceived credibility of the news. Perceived credibility was measured as the news rating, which participants gave to the items based on the 6-point Likert scale, where 1 corresponded to “fake,” 2—“most likely fake,” 3—“rather fake,” 4—“rather true,” 5—“most likely true,” 6—“true.” The credibility rating affords measures both of accuracy and user confidence in their judgment of fake and true news.

Attitude towards topics (“Attitude”). We asked respondents to express their attitude towards abortion, LGBT, and the death penalty on a 7-point Likert scale where 1—“definitely do not support,” and 7—”definitely support.” For items with positive valence we expect a positive relation between user attitude and credibility rating and for items with negative valence a negative one. Importantly, the stronger the alignment of attitude with the valence of the topic, the stronger should be the confirmation bias. Similarly, the sentiment of the comment accompanying the news items could also be consistent and discrepant with the attitude towards the topic.

Media professionalism ("MP"). We classified respondents as media professionals based on the question "Have you ever worked in media?" (No / Yes, I am working now / Yes, I worked in the past). Then participants answered a follow-up question regarding their current or last position in the media. Several respondents whose job was irrelevant to the media industry (e.g., "storekeeper" or "orchestra musician") were manually excluded from the sample. For the purposes of this study, both current (N=319) and former (N=172) media employees were included in the media professionals group.

Familiarity with the news ("Seen"). Since prior exposure could increase the perceived credibility of the news (Pennycook, Cannon & Rand, 2018), participants were asked which of the news items they had already seen before.

Verification ("Checked"). Assuming a possibility that participants could have done fact-checking while undergoing the experiment, we asked which of the news items they had verified before evaluating them.

Government support ("gsp"). As a control variable reflecting participants' political views, we have included government support, since in authoritarian regimes audience polarization can often happen based on the pro-government or pro-opposition stances (Urman, 2019).

Data analysis

Credibility ratings

Credibility ratings (i.e., ratings on Likert scales) usually serve as the dependent variable for the assessment of experimental effects of individual or message-level factors (e.g., Bryanov et al, 2022). Traditionally, ratings have been analyzed with linear mixed models to control for clustering of responses associated with random factors of user and news items using the lme4 package (Bates et al, 2015b) in the R environment for statistical computing (R Core Team, 2022). Here we use them as input for a signal-detection theory (SDT) analysis (see next section). However, we will provide some credibility-based figures for comparison with the SDT approach. In general, for preprocessing and graphics we relied on the suite of tidyverse packages (Wickham & Grolemund, 2016). Data, scripts, and figures are available in the OSF repository.

Signal detection theory (SDT)

SDT assumes that reading both fake and true news provides some amount of subjective evidence for the truth of news (irrespective of the news veracity). Obviously, on average this evidence is expected to be larger for true than fake news, but, as sketched in Figure 2a, the variance associated with them is usually large enough to cause overlap between the two distributions. Figure 2a represents the two additional classic SDT assumptions that the amount of subjective evidence is normally distributed with equal variance for both fake and true news.

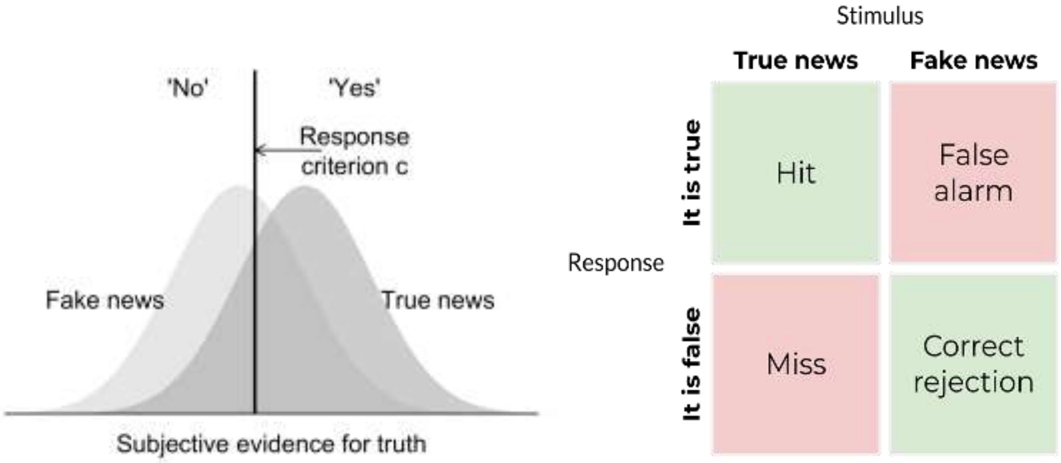


Figure 2. (a). Hypothetical distributions of subjective evidence for news items being true for fake (left) and true (right) news. Discriminability *d’* estimates the difference between the means of the two distributions; the location of criterion *c* determines whether a subject responds “not true/ fake” (evidence <= *c*) or “true” (evidence > *c*). (b) Assuming the area under each curve is 1.0, the location of the criterion *c* determines p(hit), that is of “true” response for true news items) and p(false alarm, that is “true” responses for fake news items; p(miss) is 1.0 - p(hit) and p(correct rejection) is 1.0 - p(false alarm).

With these assumptions in place, *discriminability d´* (i.e., the ability to discriminate between the two types of stimulus) is a bias-free estimate of the difference between the means of the two distributions. The smaller this distance, the lower the ability to discriminate true messages from false ones. Low values of *d’* could be due to a lack of knowledge or a very high similarity of fake and true news. Experimental manipulations, individual differences between participants (e.g., media professionalism), or differences between items can affect discriminability *d’.*

What about the response tendency to say “no — fake” or “yes — true”? SDT assumes that there must be some finite criterion value on the x-axis that determines whether the response is “fake” (i.e., when evidence is smaller than *c*) or “true” (i.e., when evidence is larger than *c*). The confidence ratings of 1 to 6 are assumed to represent six contiguous segments on the evidence axis separated by five different *c* locations.

Again, experimental manipulations, individual differences between participants (e.g., alignment of attitude and item valence), and differences between items can affect the location of the criterion *c.* For six ratings there are five locations of *c* that represent varying amounts of response bias: the further to the left, the greater the tendency to say “yes” (and vice versa to say “no”).

The vertical line at location *c* in Figure 2a is one cut through the two distributions of subjective evidence. Taking into account the veracity of items, there are four types of responses: (1) *hits* —judging a true news item as true); (2) *false alarms* —judging a fake news item as true; (3) *misses* —judging a true news item as fake; (4) *correct rejections* — judging a fake news item as fake. If we assume that the area under each curve is 1.0, then the area to the right of *c* under the “true” distribution is the probability of a hit [p(hit)] and the corresponding area under the “fake” distribution is the probability of a false alarm [p(fa)]; the complements of these probabilities are p(miss) and p(correct rejection), respectively. The same statistics can be computed for the other criterion locations. With each "move" from one criterion location to the next, we decrease the hit and false alarm rate, that is we decrease the bias to say "yes" at the same discriminability.

Based on the probit-transformation of hit and false-alarm probabilities (i.e., their transformation into z-scores), the two estimates are computed as:

(1) *d'* = z(hit rate) - z(fa rate)

(2) *c* = [z(hit rate) + z(fa rate)]/2

c is the average of the probits of hit and false-alarm rates and *d'* is the difference between them. Although hit and false-alarm rates may be correlated, their transformation to average and difference of z-scores renders them mathematically orthogonal (uncorrelated), because with this transformation they represent the two principal components of the original rates.

Cumulative link mixed model

Porshnev et al. (2022) describe how SDT parameters and effects on SDT parameters can be estimated from the four response rates with probit regressions. Each of the four response rates can be understood as a conditional probability for a response category ("yes" or "no") given the veracity of the presented news item, Pr(response | veracity). As such, we can also write Equations (1) and (2) as:

(3) *d'* = z[Pr("yes" | true news)] - z[Pr("yes" | fake news)]

(4) *c* = { z[Pr("yes" | true news)] + z[Pr("yes" | fake news)] } / 2

This allows us to formulate the probability for a "yes" response as a function of parameters *d'* and *c*, and the veracity *X* of the news item in a single probit model,

(5) Pr("yes" | *X*) = Φ[-*c* + *X d'*] ,

where the cumulative distribution function Φ of the normal distribution is the inverse of the z-transformation and the predictor *X* for veracity is either +½ for true news or –½ for fake news.

It follows that, with the appropriate specification, discriminability *d'* and response bias *c* as well as experimental effects and their interactions on them can be estimated with a cumulative link model, available as the *clm* function in the *ordinal* package in R (Christensen, 2019). The cumulative link model estimates these fixed effects with probit regressions for the five criterion locations. The *clm* function assumes that responses are independent. Obviously with twelve responses per subject and hundreds of responses per item this assumption is not met. However, the cumulative link mixed model *clmm* function, also available in the *ordinal* package, allows the specification of crossed random factors for subjects and items. Moreover, when supported by the data, variance components (VCs) and correlation parameters (CPs) can be estimated for within-subject and within-item effects (see Porshnev et al, 2022, for details and tutorial introducing the *SDTvis* package in R).

Tests of hypotheses 1 to 4 were available with a *clmm* specifying veracity (2) x valence (2) x comment (3) x media professionalism (2) as nested within the three topics. We also included whether the news was rated as seen before and whether its veracity was checked prior to the ratings. Finally, three variables coding participants' attitude towards each of the topics, their overall support of the government, and their age served as covariates.

Sequence of model tests and model selection

We tested the four sets of hypotheses as nested under a topic, meaning we obtain a separate set of test statistics for each topic, but estimated in an integrated CLMM. Aside from the topic, the core CLMM comprises the independent variables veracity and valence of news and individual differences in attitudes towards the three topics. This core model affords a test of Hypothesis 1 for the three topics. In a second step, the between-subject variable media professionalism and its interactions with valence and user attitude were added to test Hypotheses 2 and 3. In a third step, we tested the amplification of confirmation bias due to comments accompanying the news (Hypothesis 4). In a fourth step, we test the effects of covariates, that is users' belief of having seen the news and users' checking the veracity of the news prior to providing the credibility rating. In the final step, we add two covariates for the effects of general support of government and age.

The steps yield a sequence of nested CLMMs for which hypotheses can be tested with likelihood-ratio tests (LRTs). At each step, we can see whether effects at an earlier stage are still significant when significant new variables are added to the model. The random-effect structure of the CLMMs comprises two random factors yielding estimates of subject-related variance components for discriminability *d'* and response bias *c* and, because veracity is a between-item factor, only an item-related variance component for response bias *c.*

Results

Descriptive statistics

Table 1 summarizes means and standard deviations of covariates and credibility ratings. Media professionals have a more positive attitude towards abortion and LGBT and a more negative attitude towards the death penalty. Overall, attitudes are most positive towards abortion, followed by LGBT, and least positive for the death penality.

Table 1. Descriptive statistics

Topic	MP	N	Age		Gov support		Attitude		Credibility	
			M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Abortion	no	1455	38.3	*10.1*	3.26	*1.81*	5.35	*1.89*	3.58	*0.97*
	yes	491	36.2	*10.2*	3.09	*1.71*	5.73	*1.79*	3.70	*0.94*
LGBT	no	1455	38.3	*10.1*	3.26	*1.81*	3.60	*2.35*	3.64	*0.98*
	yes	491	36.2	*10.2*	3.09	*1.71*	4.57	*2.26*	3.58	*1.01*

Topic	MP	N	Age		Gov support		Attitude		Credibility	
D penalty	no	1455	38.3	*10.1*	3.26	*1.81*	3.89	*2.17*	3.63	*0.98*
	yes	491	36.2	*10.2*	3.09	*1.71*	3.24	*2.12*	3.60	*0.91*

Note. MP = media professional. Ratings of government support and attitude range from 1 to 7; ratings of credibility from 1 to 6.

Table 2 presents correlations for participant variables. Attitudes towards the three topics correlate largely as expected. They are most positive for abortion and LGBT and most negative for LGBT with age, goverment support, and death penalty for both groups.

Table 2. Correlations between individual-difference variables and attitudes towards topics

	Age	Gov supp	Abortion	LGBT	Penalty
Age	1.00	0.18	-0.13	**-0.30**	0.07
Gov support	0.14	1.00	-0.27	**-0.48**	0.32
Abortion	-0.15	-0.21	1.00	**0.52**	-0.09
LGBT	**-0.31**	**-0.38**	**0.42**	1.00	**-0.37**
Death penalty	0.05	0.20	-0.06	**-0.37**	1.00

Note. Correlations above diagonal are for media professionals (N=491) and below for non-professionals (N=1455); bold *r* > .30.

Incremental model tests

In the sequence of model tests, the core CLMM comprised effects of valence, attitude, and their interaction as nested under the three topics. Adding media professionalism (MP) main effects significantly improved the goodness of fit; $\chi^2(6) = 15.2$, p = 0.019. However, adding interactions of MP with valence and attitude did not improve the goodness of fit; $\chi^2(18) = 16.5$, p = 0.559. Thus, there is evidence that MP plays a role (see below), but there is no evidence for Hypothesis 3 that media professionals are less susceptible to confirmation bias than ordinary social media users.

Adding the main effect of comment sentiment and its interactions also did not lead to any significant improvement; main effects: $\chi^2(12) = 0.39$ p= 1.00, interactions: $\chi^2(36) = 23.9$ p = 0.94. Thus, this experimental manipulation and possible moderations by MP were not significant. Hypothesis 4a and 4b did not receive any support in this experiment.

Finally, both pairs of covariates (along with some interactions) were significant; familiarity with the news (“seen”) and checks of the veracity prior to the rating: $\chi^2(24) = 286.1$ p<.001; government support and age: $\chi^2(24) = 91.0$, p<.01.

The fitted model objects are provided in the OSF repository. Table 3 lists all significant effects (p < .05) of the final CLMM. With two exceptions related to media professionalism (MP), they were also significant in the three simpler models.

Table 3. Significant fixed-effect estimates nested under topic for final CLMM

	Abortion	*LGBT*	*Death penalty*
Response bias c			
Valence	0.326 (0.069)		
Attitude	0.016 (0.007)		
Valence x attitude	**0.058 (0.007)**	**0.047 (0.006)**	**0.031 (0.006)**
Seen	**0.289 (0.056)**	**0.334 (0.057)**	**0.239 (0.060)**
Valence x checked	-0.177 (0.076)	0.186 (0.074)	
MP x checked			0.169 (0.079)
Gov support (gsp)	-0.019 (0.008)		
Age		-0.003 (0.001)	0.004 (0.001)
Discriminability d'			
Grand Mean of d'		0.140 (0.069)	0.172 (0.069)
Valence			0.162 (0.068)
Attitude			-0.022 (0.006)
Valence x attitude	-0.014 (0.007)		
MP	*CLMM 2*	*CLMM 2*	**-0.034 (0.015)**
Checked	**0.382 (0.077)**	**0.501 (0.077)**	**0.572 (0.078)**
MP x checked	**0.270 (0.077)**	**0.277 (0.077)**	**0.136 (0.078)+**
Valence x seen	-0.165 (0.051)		
Gov support (gsp)			-0.017 (0.007)

	Abortion	*LGBT*	*Death penalty*
Valence x gsp	0.024 (0.007)		
Valence x age	0.003 (0.001)	-0.006 (0.001)	

Note. MP = media professional; bold = consistent across topics; *CLMM2 = effect significant in CLMM without Check covariate.* + : p < .10. Only significant effects (p < .05); a complete list in *Supplement*.

Confirmation bias: valence of news x user attitude

There is a strong pattern of confirmation bias for news evaluation. As demonstrated in Figure 3a (top row), criterion *c*, which reflects the response bias, shifts depending on the news valence and an attitude towards a topic. According to our data, a more positive attitude about the topic increases the criterion for news items with positive valence, whereas a negative attitude increases the criterion for news items with negative valence. Thus, if the news is aligned with participants' views, it is perceived as more credible. The valence x attitude interaction is observed for all three topics (p < .001; see Table 3).

Yet, discriminability *d'* (Figure 3a, bottom row) does not change in a systematic way depending on the news congruence for abortion and LGBT topics. The only significant effect is observed for the death penalty topic: participants in favor of the death penalty generally discriminate news about this issue worse than those who are against it, regardless of the news valence (p < .05).

The original credibility rating (shown in Figure 3b) aligns much more with response bias than with discriminability. Thus, the difference in the perceived news credibility is primarily explained by confirmation bias rather than by the discriminability of the items, which fully confirms Hypothesis 1.

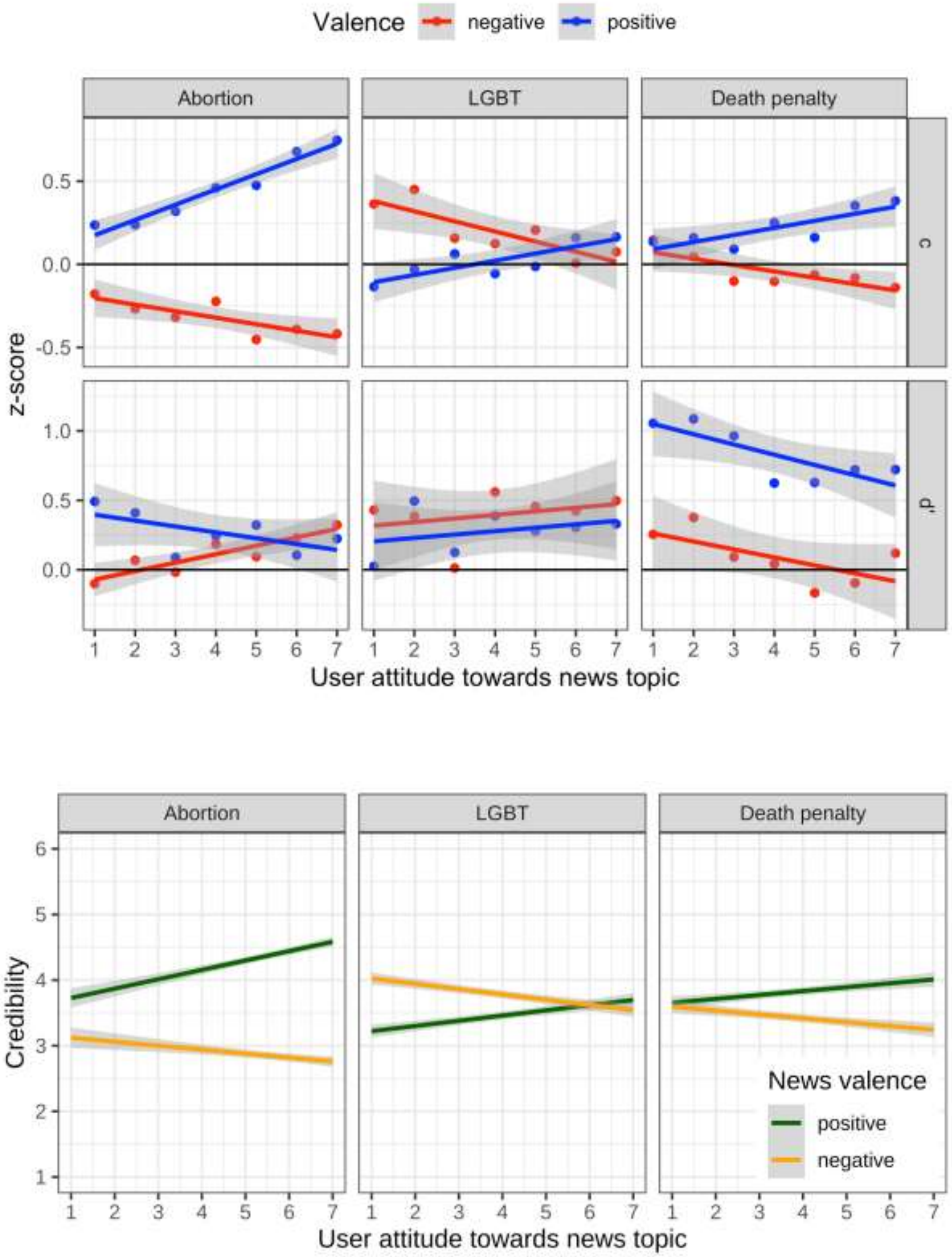
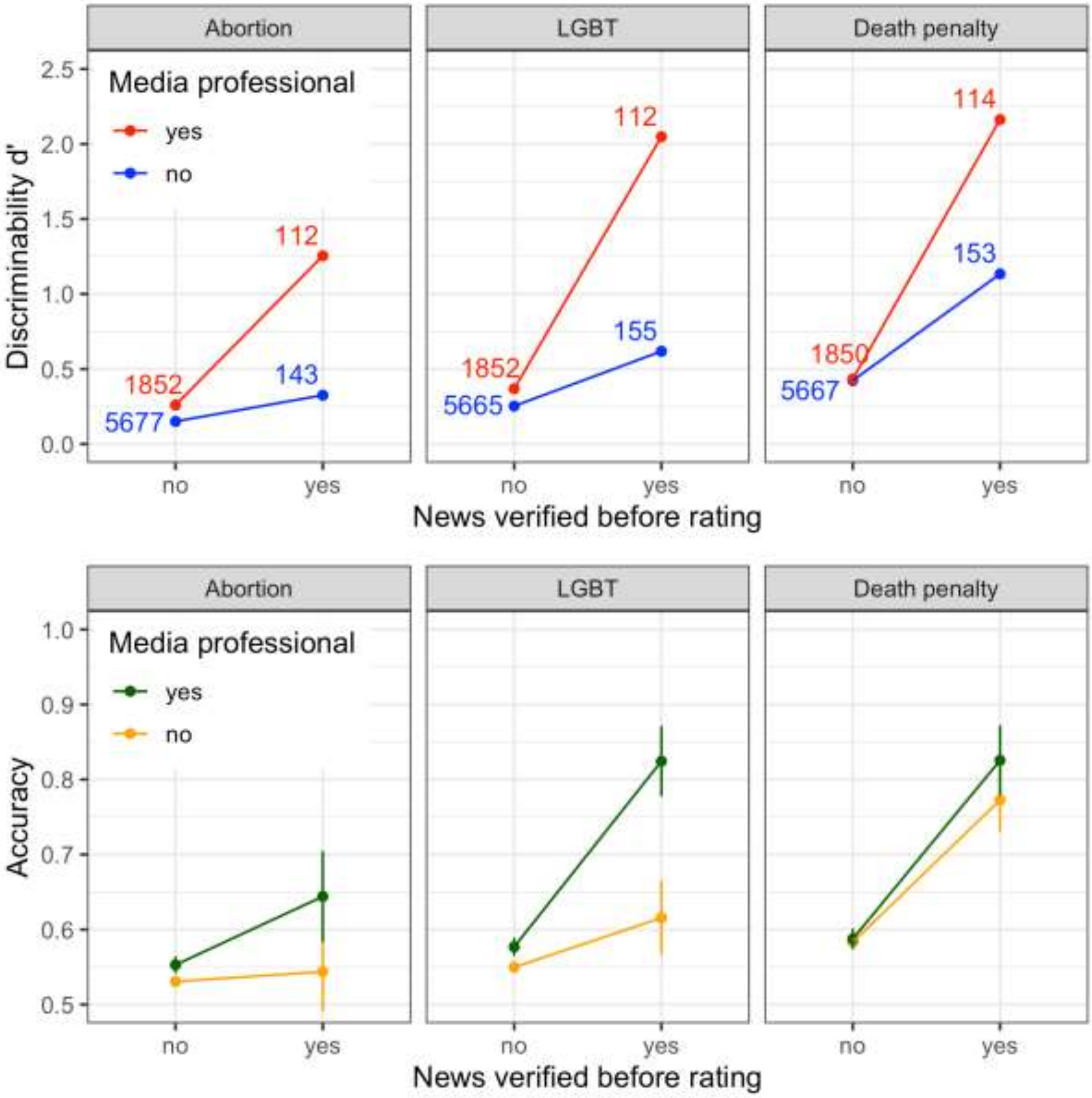


Figure 3. (a) Criterion *c* (top) and discriminability *d'* as a function of user attitude towards topic and valence for the three topics. Observed means (dots) are fitted with a linear regression; error band is 95% confidence interval for regression. (b) regressions for credibility ratings from which SDT indices were computed.

Media professionalism: discriminability and confirmation bias

Media professionals had significantly better discriminability *d'* for news about abortions and LGBT, but not for the death penalty, when it was added as a main effect to the core CLMM in step 2. When, in step 4 (see below), we added the covariate coding news verification prior to the rating ("checked") and its interaction with media professionalism, the main effects on discriminability *d'* were no longer significant, but the interaction with "checked" was (see Table 3). In addition, the main effect of media professionalism was now significant for the death-penalty topic. The interactions are shown for the three topics in *Figure 4a.* When media professionals do not check the news, they are not too different from users without professional media experience, but when they do check, they greatly outperform them. It means that checking per se is not enough, it also depends on the verification quality. Thus, we consider Hypothesis 2 to be supported when media professionals apply their fact-checking skills.



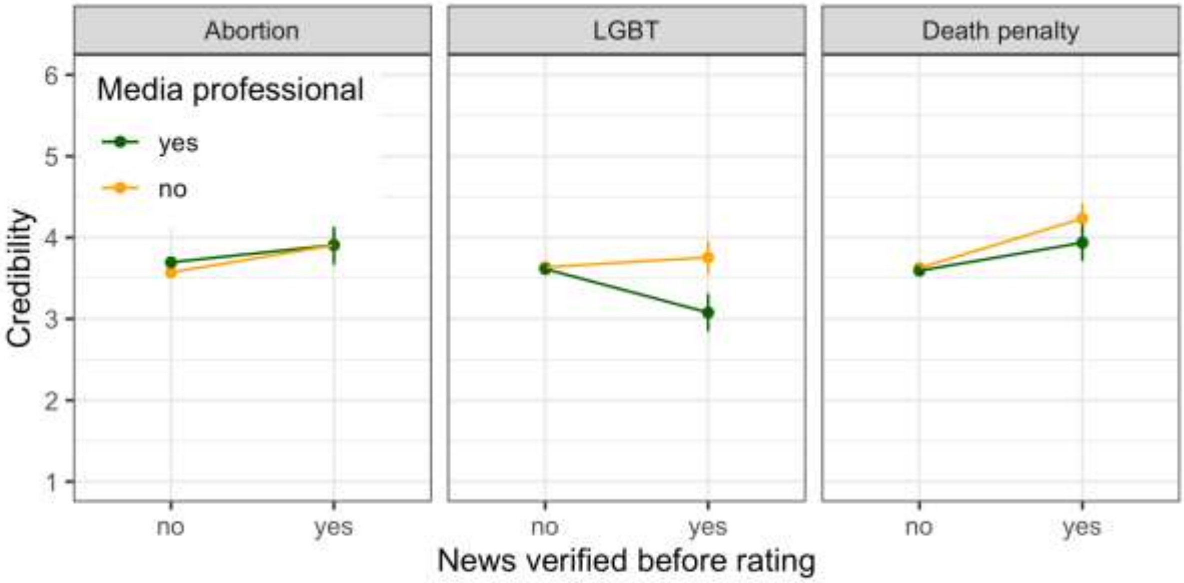


Figure 4. (a) Discriminability *d'* when news items were not or were verified before rating news by media professionals and normal users; there were no significant effects on response bias *c.* (b) The same profile of means for accuracy. (c) The same profile of means for credibility ratings.

Moreover, it appears that the advantage of media professionals stems from their greater willingness to verify the news. Although the overall percentage of news that was checked was very small (3.4%), this percentage was more than twice for media professionals (5.7%) than for normal users (2.6%). News verification occurred about equally often for true and fake news for both groups. As shown in Figure 4, even though about three times as many ordinary users as media professionals participated, they did not differ by very much in the absolute number of checked news.

Finally, as shown in Figure 4b, discriminability *d'* aligns reasonably, but not identically, with accuracy, that is with p(hit)+p(correct rejection). The credibility rating would not have allowed us to uncover this result (Figure 4c).

As already described in *Incremental model tests*, there is no evidence that media professionals differ in confirmation bias when compared to other users. None of the interactions involving valence and attitude towards the topic were significant. Thus, Hypothesis 3 was not supported.

Confirmation bias: comment sentiment x user attitude

As described in the *Incremental model tests*, when we added the comment-sentiment manipulation (positive, neutral, negative) to the model. Contrary to Hypotheses 4a and 4b, we have no evidence for an effect of the news comment on the discriminability or response bias — neither as a main effect, nor in interaction with the valence of the news or the user attitude towards the topics, nor MP; for all LRTs the change in χ^2 was always smaller than the degrees of freedom. Thus, Hypotheses 4a and 4b are clearly rejected for this study.

Familiarity with the news and news verification prior to rating

In the fourth step, we included two control covariates: perceived familiarity with the news ("seen") and verification prior to rating ("checked"). Only 7.9% of all news items were rated as seen prior to the experiment with only a slight advantage for true news (52%) and, as mentioned above, only 3.4% of news items were rated as verified prior to entering the rating with an even smaller advantage for fake news (51%). Despite the small number of observations associated with these covariates, they had the largest effects.

News (believed) seen. News elicits a strong positive response bias if users believe that they saw the news before or actually saw it. This effect was significant for all three topics. Conversely, there was no systematic effect direction associated with discriminability *d'* (see Table 3). We will return to the implications of this result in the *Conclusion*, but likely participants responded more to familiar aspects of the news and less to details that rendered it fake.

News checked. A reported act of verification during the experiment, not unexpectedly, increases discriminability *d'* for all three topics (see Table 3). If a participant checked the news while going through the experiment, this increased their accuracy in news discernment, but as shown in Figure 3, the effect is strongly amplified by media professionalism.

General covariates: support of government and age

In the final step, we added users' government support and age to the model. They significantly improved the goodness of fit, but the effects were scattered across topics and interactions (see Table 3). While plausible interpretations of these effects can be offered, for many of them, most of them are post-hoc and, given that they do not replicate across the three topics, serve primarily heuristic purposes for follow-up experiments and analysis. Figures of the valence x government support and valence x age interactions for the three topics are provided in the OSF repository.

Conclusion

Using an online experiment among Russian social media users and media professionals, this study examines the role of media competence and confirmation bias in the evaluation of news credibility. Although there were other experiments studying how journalists judge various kinds of information (e.g., McGregor & Molyneux, 2020; Graves, Nyhan & Reifler, 2016), to our knowledge, this is the first attempt to test how media professionals discriminate between true and fake news. This study reveals several main findings:

First, there is a very strong effect of confirmation bias regardless of the news topic and media professionalism. In particular, participants are more likely to believe news articles that are aligned with their beliefs than contradicting ones, which confirms the results of earlier studies (e.g., Kim & Dennis, 2019). With the help of signal detection theory, we clearly demonstrate that this effect is primarily explained by the response bias (tendency to say "yes" to the stimulus, or to judge news as true) rather than by the difference in news discriminability. Thus, humans' propensity to believe information consistent with their views can override professional intuition. Yet, according to our results, user comments—either supporting or contradicting the reader's beliefs—do not in any way affect the response bias. One of the possible explanations for this is that comments generally matter less for the news evaluation than existing attitudes (Steinfeld, Samuel-Azran, & Lev-On, 2016).

Second, media professionals are indeed more accurate in news discrimination than usual social media news consumers (although not more immune to confirmation bias). However, their ability to discriminate the news is explained not by the inherent propensity to detect falsehoods but rather by more advanced fact-checking skills. In our study, we did not force participants to fact-check the news but we assumed they could do it during the task so we asked which of the articles they verified. For unverified news articles, the discriminability for both media professionals and ordinary users was nearly on the same level. However, for fact-checked news, media professionals showed a much higher level of discriminability than other participants. Thus, our study suggests that media professionals are not less susceptible to disinformation or confirmation bias but they are more skilled in fact-checking and more inclined to perform it, which, in the end, helps them to better recognize fake news.

Third, we find that news articles checked during the experiment are generally judged more accurately (largely irrespective of response bias), although, as already mentioned, this effect is much higher for media professionals. Meanwhile, news articles rated as familiar (seen before the experiment) are perceived as more credible (largely irrespective of accuracy). This goes in line with the results of prior research, which suggests that fake news is perceived as more credible if a person has been previously exposed to it (Pennycook, Cannon, & Rand, 2018). Thus, media literacy skills (in a form of active fact-checking) have the potential to improve people's ability to recognize disinformation, whereas repeated exposure to particular stories or narratives can, on the contrary, increase belief in falsehoods.

Most of the mis- and disinformation research focuses on the influence of individual or message characteristics on people's ability to recognize fake news or their likelihood to believe it. Yet, these studies mostly ignore the way people arrive at their judgments, namely—how and if they verify the information they consume. By giving participants a difficult task of discriminating between very similar true news and fake news, this study shows that the actual act of fact-checking is the only factor positively affecting accuracy in news discernment even for media professionals. Thus, we contribute to the existing literature on fake news and journalism by demonstrating that it is developing fact-checking skills that actually makes individuals more resistant to mis- and disinformation. Furthermore, we suggest that instead of solely trying to answer *why* people believe fake news, research should also look into *how* they evaluate it.

Another contribution of this study is the implementation of the SDT approach to fake news research. To our knowledge, there is only one study (Batailler et al, 2022), developed in parallel to our research and available online since July 2021, that provided an introductory overview of using SDT in the field of fake news research. The authors illustrate the approach with secondary analyses of true and fake news headlines. However, since there were no ratings of observers' confidence available for these data, the SDT parameters could only be computed descriptively from hit and false-alarm rates, but not estimated in a cumulative link mixed model (CLMM). Furthermore, critical SDT assumptions (e.g., normal distribution with an equal variance of subjective evidence for true and fake news items) could not be tested. Indeed, the authors acknowledge these limitations and call exactly for the type of experiment we report here.

Within a user, discriminability and response bias are two uncorrelated SDT features (i.e., principal components), they are both mathematically and conceptually independent of each other. Thus, the response of a user is determined (to a degree varying from null to complete) by their discriminability and their response bias independently of each other. This means that individuals with the same level of discriminability may have different response biases, either positive or negative. Moreover, both discriminability and response bias can be affected independently or jointly by experimental manipulations, individual differences of the users, differences between the items, and the interactions of the two latter (e.g. user attitudes to a topic and message valence in regards to this topic). Importantly, correlations between sensitivity and response bias may emerge when we move from within-subject to between-subject effects. For example, high discriminability may "protect" against bias, that is correlate negatively with bias, despite the fact that discriminability and response bias are uncorrelated within subjects. SDT offers a relatively simple method to calculate discriminability and response bias based on the number of "hits" (instances of true news correctly recognized as true) and "false alarms" (instances of fake news falsely judged to be true).

Signal detection theory was developed to remove the contribution of response bias to perceptual sensitivity in psychophysics experiments. These experiments often comprised a handful of subjects performing thousands of judgments about the presence/absence of a stimulus in noise. Our adoption of signal detection theory for research on political communication reverses this basic design feature: in a web experiment, we collect data from thousands of subjects who rate only a very limited number of messages (i.e., twelve in our study). We consider it reassuring that despite such a dramatic change in experimental design expected and unexpected but quite plausible effects on response bias and discriminability could be recovered with this methodology.

Limitations

Nonetheless, we must address some limitations of this study. First of all, there is no objective measurement of media literacy or media competence in our experiment. As self-reported media professionalism might be in fact not equivalent to objectively measured media literacy, this may have led to the same levels of susceptibility to fakes between professionals and ordinary users in our study. In particular, not all media professionals have first-hand experience in news production and/or work in news organizations (for example, independent bloggers also fell in the category of media professionals). Furthermore, since participation in the study was anonymous, we did not ask for the names of media organizations our participants worked for and, thus, could not evaluate the quality of their journalistic work.

Another limitation of our study might be at the core of the null effect of the sentiment of the comment accompanying the news. It might be explained not necessarily by the unimportance of comments but also by the experimental design: since comments did not include any factual information useful for news evaluation, participants could have started skipping them and paying more attention to a news item itself. Whether this is true only in a web-experimental setting or in general, remains to be determined. One way to discriminate between the effect of comment unimportance and inattention to it is to force readers to read the comment before they are presented with the news and/or control whether they do it with an eye tracker.

Another limitation concerns the false news items. Since they were constructed specifically for the needs of this study, these items might have been less easy to recognize than "real" fake news. In particular, these articles were written in a similar manner to true news and, therefore, did not have some of the common fake news features such as emotionality, capitalization, or punctuation mistakes (as described in Damstra et al, 2021), which could serve as cues to the participants. This construction aspect allowed us to clearly trace the influence of response bias (including confirmation bias) on the news evaluation but might have made fake news recognition more difficult. Yet, this way our study also shows how users may judge thoroughly constructed disinformation, which undoubtedly also exists in the media environment.

The study has several implications for future research. First, for a better evaluation of the role of media professionalism in fake news perception, it is necessary to include media literacy scales that could measure actual knowledge among media professionals and ordinary users. Second, among media professionals, a narrower group of professional journalists (e.g., reporters and editors working for the news organization) needs to be explored in more detail to determine the factors, if any, that make them different from others in terms of susceptibility to disinformation. Finally, since verification (especially, if done professionally) proved to have a positive effect on news discrimination, it is crucial to further investigate actual fact-checking practices of Internet users and their comparative efficiency.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Altay, S., Berriche, M., & Acerbi, A. (n.d.). *Misinformation on Misinformation: Conceptual and Methodological Challenges*. 1–28. https://doi.org/10.31234/osf.io/edqc8

Amazeen, M. A., & Bucy, E. P. (2019). Conferring Resistance to Digital Disinformation: The Inoculating Influence of Procedural News Knowledge. *Journal of Broadcasting & Electronic Media*, *63*(3), 415–432. https://doi.org/10.1080/08838151.2019.1653101

Aufderheide, P. (1993). *Media Literacy: A Report of The National Leadership Conference on Media Literacy*.

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A Signal Detection Approach to Understanding the Identification of Fake News. *Perspectives on Psychological Science*, *17*(1), 78 –98. https://doi.org/10.1177/1745691620986135

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2015a). Parsimonious mixed models. *arXiv preprint*, *1506.04967*, 1–27. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Brandtzaeg, P. B., Lüders, M., Spangenberg, J., Rath-Wiggins, L., & Følstad, A. (2016). Emerging Journalistic Verification Practices Concerning Social Media. *Journalism Practice*, *10*(3), 323–342. https://doi.org/10.1080/17512786.2015.1020331

Bryanov, K., Kliegl, R., Koltsova, O., Porshnev, A., Lokot, T., Miltsov, A., Pashakhin, S., Sinyavskaya, Y., Terpilovskii, M., & Vziatysheva, V. (2022). What Drives Perceptions of Foreign News Coverage Credibility? A Cross-National Experiment Including Kazakhstan, Russia, and Ukraine. *(Under review).*

Bryanov, K., & Vziatysheva, V. (2021). Determinants of individuals' belief in fake news: A scoping review determinants of belief in fake news. *PLoS ONE 16(6):*, *16*(6), 1–25. https://doi.org/10.1371/journal.pone.0253717

Carrieri, V., Madio, L., & Principe, F. (2019). Vaccine hesitancy and (fake) news: Quasi - experimental evidence from Italy. *Health Economics*, *28*, 1377–1382. https://doi.org/10.1002/hec.3937

Čavojová, V., Šrol, J., & Adamus, M. (2018). My point is valid, yours is not: myside bias in reasoning about abortion. *Journal of Cognitive Psychology*, *30*(7), 656–669. https://doi.org/10.1080/20445911.2018.1518961

Christensen, R. H. B. (2019). *ordinal:* Regression models for ordinal data [R package version 2019.12-10]. https://CRAN.R-project.org/package=ordinal

Council of the European Union. (2020). *Council conclusions on media literacy in an ever-changing world.* Retrieved from https://www.consilium.europa.eu/media/44117/st08274-en20.pdf

Damstra, A., Boomgaarden, H. G., Broda, E., Lindgren, E., Strömbäck, J., Tsfati, Y., & Vliegenthart, R. (2021). What Does Fake Look Like? A Review of the Literature on Intentional Deception in the News and on Social Media. *Journalism Studies*, 1–17. https://doi.org/10.1080/1461670X.2021.1979423

Deuze, M. (2005). What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism*, *6*(4), 442–464. https://doi.org/10.1177/1464884905056815

European Commission. (2007). *A European approach to media literacy in the digital environment. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52007DC0833&from=el*

Graves, L., Nyhan, B., & Reifler, J. (2016). Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking. *Journal of Communication ISSN*, *66*, 102–138. https://doi.org/10.1111/jcom.12198

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*.

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

Himma-Kadakas, M. (2017). Alternative facts and fake news entering journalistic content production cycle. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, Vol. 9, pp. 25–40. https://doi.org/10.5130/ccs.v9i2.5469

Inglehart, R. F. (2018). *Cultural Evolution: People's Motivations are Changing, and Reshaping the World*. https://doi.org/10.1017/9781108613880

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does Media Literacy Help Identification of Fake News? Information Literacy Helps, but Other Literacies Don't. *American Behavioral Scientist*, *65*(2), 371 –388. https://doi.org/10.1177/0002764219869406

Kim, A., & Dennis, A. R. (2019). Says who? The effects of presentation format and source rating on fake news in social media. *MIS Quarterly*, *43*(3), 1025–1039. https://doi.org/10.25300/MISQ/2019/15188

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation Bias in Online Searches: Impacts of Selective Exposure Before an Election on Political Attitude Strength and Shifts. *Journal of Computer-Mediated Communication*, *20*, 171–187. https://doi.org/10.1111/jcc4.12105

Knobloch-Westerwick, S., & Meng, J. (2009). Looking the Other Way: Selective Exposure to Attitude-Consistent and Counterattitudinal Political Information. *Communication Research*, *36*(3), 436–448. https://doi.org/10.1177/0093650209333030

Knobloch-Westerwick, S., Mothes, C., Johnson, B. K., Westerwick, A., & Donsbach, W. (2015). Political Online Information Searching in Germany and the United States: Confirmation Bias, Source Credibility, and Attitude Impacts. *Journal of Communication*, *65*, 489–511. https://doi.org/10.1111/jcom.12154

Köbis, N. C., Doležalova, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience*, *24*, *103364*. https://doi.org/10.1016/j.isci.2021.103364

Lazer, D., Baum, M., Benkler, Y., Berinsky, A., Greenhill, K., Menczer, F., … Zittrain, J. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, *5*, 337–348. https://doi.org/10.1038/s41562-021-01056-1

McGregor, S. C., & Molyneux, L. (2020). Twitter’s influence on news judgment: An experiment among journalists. *Journalism*, *21*(5), 597 –613. https://doi.org/10.1177/1464884918802975

Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports*, *12*(6008). https://doi.org/10.1038/s41598-022-08437-0

Moravec, P. L., Minas, R. K., & Dennis, A. R. (2019). Fake News on Social Media: People Believe What They Want to Believe When it Makes No Sense At All. *Mis Quarterly*, *43*(4), 1343–1360. https://doi.org/10.25300/MISQ/2019/15505

Newman, N., Fletcher, R., Schulz, A., Andı, S., & Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

PEN America. (2021). *The Impact of Disinformation on Journalism: Online Survey of Journalists*. Retrieved from https://pen.org/report/hard-news-journalists-and-the-threat-of-disinformation/

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior Exposure Increases Perceived Accuracy of Fake News. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Porshnev, A., Rabe, M., Terpilovskii, M., & Kliegl, R. (2022). Sensitivity and Bias in the Discrimination of True and Fake News: A Signal Detection Theory Approach (in preparation)

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo Chambers on Facebook. *SSRN Electronic Journal*, 1–15. https://doi.org/10.2139/ssrn.2795110

R Core Team. (2022). R: A language and environment for statistical computing. https://www.r-project.org/

Saldaña, M., & Vu, H. T. (2021). You Are Fake News! Factors Impacting Journalists' Debunking Behaviors on Social Media. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2021.2004554

Silverman, C. (2015). *Lies, Damn Lies and Viral Content*. https://doi.org/10.7916/D8Q81RHH

Steinfeld, N., Samuel-Azran, T., & Lev-On, A. (2016). User comments and public opinion: Findings from an eye-tracking experiment. *Computers in Human Behavior*, *61*, 63–72. https://doi.org/10.1016/j.chb.2016.03.004

Taber, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation ofPolitical Beliefs. *American Journal OfPolitical Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*(6), 401–409. https://doi.org/10.1037/h0058700

Taub, H. A. (1965). Effects of differential value on recall of visual symbols. *Journal of Experimental Psychology*, *69*(2), 135–143. https://doi.org/10.1037/h0021591

Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, *44*(2), 157–173. https://doi.org/10.1080/23808985.2020.1759443

Urman, A. (2019). News Consumption of Russian Vkontakte Users: Polarization and News Avoidance. *International Journal of Communication*, *13*, 5158–5182.

Vziatysheva, V., Sinyavskaya, Y., Porshnev, A., Terpilovskii, M., Koltcov, S., & Bryanov, K. (2021). Testing Users' Ability to Recognize Fake News in Three Countries. An Experimental Perspective. In G. Meiselwitz (Ed.), *Social Computing and Social Media: Experience Design and Social Network Analysis. HCII 2021. Lecture Notes in Computer Science* (pp. 370–390). https://doi.org/10.1007/978-3-030-77626-8

Westerwick, A., Johnson, B. K., & Knobloch-Westerwick, S. (2017). Confirmation biases in selective exposure to political online information: Source bias vs . content bias. *Communication Monographs*, *84*(3), 343–364. https://doi.org/10.1080/03637751.2016.1272761

Wickham, H., & Grolemund, G. (2016). *R for data science*. Beijing: O'Reilly.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 201–233. https://doi.org/10.1037/xlm0000732

Zimmermann, F., & Kohring, M. (2020). Mistrust, Disinforming News, and Vote Choice: A Panel Survey on the Origins and Consequences of Believing Disinformation in the 2017 German Parliamentary Election. *Political Communication*, *37*, 215–237. https://doi.org/10.1080/10584609.2019.1686095

Acknowledgements

This article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University)

We are grateful to Maximilian M. Rabe for advice and comments.

Data and scripts are available at the OSF repository https://osf.io/3tmk7/.

Address for correspondence: Victoria Vziatysheva (victoria.vztshv@gmail.com) or Reinhold Kliegl (reinhold.kliegl@uni-potsdam.de)

Appendix

Examples of the stimulus material: news items and comments (translation from Russian)

Note: true news items were taken from the news media outlets; fake news items were constructed for the purposes of this research. Comments (for both fake and true news) were found on social media in discussions of the same or similar news. "Positive" and "negative" in cases of both news valence and comment sentiment reflect the stance toward the topic: e.g., a negative comment regarding the news about abortions means that it an author expresses a negative attitude onwards abortions and not towards the news item. The examples below represent part of the sample that consisted of 24 news items with a unique pair of comments for each of them.

News item 1

Topic: Abortions
Veracity: True
Valence: Negative

At the initiative of the Russian Orthodo Church, the majority of the state medical institutions in the Yaroslavl region stoped conducting planned abortions on January 11. "The action is dedicated to the memory of the Bethlehem babies killed by King Herod, who wanted to destroy the God-child," the Metropolis said in a press release.

Positive comment: Bigotry. Shouldn't all churches be closed in memory of the lives destroyed by the Inquisition?
Negative comment: If only they did it all over the country. At least one day without abortions ..

News item 2

Topic: LGBT
Veracity: True
Valence: Positive

Avanti West Coast rail company has launched the first Pride train in the UK, which will run routes between Euston and Manchester. The train is painted in the colors of the LGBT flag and is run by all LGBT crew. During the trip, passengers will be able to read queer literature, see paintings by LGBT artists, and learn facts about the movement during the onboard announcements.

Positive comment: This is an important step to increase the visibility of LGBT people. There is too much heteronormativity and cissexism in the world. Well done Brits!

Negative comment: Thank God I live in Russia, where people have adequate family values and nobody imposes this disgusting ideology on anyone.

News item 3

Topic: LGBT
Veracity: Fake
Valence: Negative

In the Czech Republic, a Russian expat opened a bar "for straight people" as a response to the growing number of gay bars and clubs. The bar “Straight Place” is located in the city of Ostrava. Local news media outlets report that the founder may be associated with the movement called “Saw”, which attacks LGBT people.

Positive comment: All people are equal, everyone is worthy of love and respect. People have the right to live like they want and with whom they want and not hide it.
Negative comment: That’s a real man, do not have other words. LGBT is evil.

News item 4

Topic: Death penalty
Veracity: Fake
Valence: Positive

The Council of Ministers of the Republic of Bulgaria will discuss a law allowing the death penalty for particularly serious crimes. New legislation has been initiated by the Deputy Prime Minister and Minister of Defense Krasimir Karakachanov. In his opinion, serial killers, pedophiles, and terrorists deserve only capital punishment.

Positive comment: I am for the death penalty. Tired of this mess. But only after a thorough check of the guilt of the defendant .. and he must be judged not by a “troika”[1] but by a jury.
Negative comment: How many innocent people can suffer?! History knows numerous errors of justice all over the world.

[1] “Troika” is a Russian word, which in this case refers to NKVD troika, or a special commission of three officials who issued sentences after a simplified procedure and without a proper trial during the Stalin times.