Confirmation bias and users' comments in fake news detection accuracy and credibility perception: an eye tracking approach

Olessia Koltsova[1], Elena Artemenko[1*], Maksim Terpilowskii[1], Taisiia Ulianova[2,3].

[1] National Research University Higher School of Economics (HSE University), Laboratory for Social and Cognitive Informatics, Saint Petersburg, 192148, Russia

[2]University of Potsdam, Potsdam, 14469, Germany

[3]Max Planck Institute for Human Development, Berlin, 14195, Germany

[*]**Corresponding author:**
Elena Artemenko, Laboratory for Social and Cognitive Informatics, National Research University Higher School of Economics (HSE University), 55/2 Sedova, St., Petersburg, Russia, 192148. Telephone: +7 (812) 644-59-10 ext. 61312
Email: nekrasovaed@gmail.com

Abstract

Confirmation bias (CB) - a tendency to ascribe higher credibility to messages consistent with news consumers' beliefs or desires - has been widely studied as potentially damaging for the accuracy of true and false message discernment. However, the actual effect of experimentally measured CB on fake news detection accuracy has not been directly tested. Neither there exists conclusive evidence on the role of social cues, such as comments, for CB reinforcement or mitigation, although news messages are being increasingly spread through platforms allowing such cues to proliferate. To cover these gaps, we conduct a lab experiment with participants reading sets of commented and uncommented news on an emulated social media interface (N=50, Nobservations =1800). Eye movement recording is used to control whether users notice comments. We find that CB exists and indeed affects the accuracy of false and true message detection negatively, while attitude strength does not. Although most comments are read by participants, commentary valence has no effect on message believability, thus neither reducing nor strengthening CB, however, comment presence does play a role. This suggests that people may find it easier to suppress CB when nudging is presented as a separate comment rather than within a news text.

Keywords

eye tracking, fake news, confirmation bias, misinformation, perceived credibility, fake news detection accuracy

1. Introduction

As information technologies are simplifying dissemination of messages, both true and false, it has become apparent that incorrect evaluation of message veracity by information recipients may have negative social consequences (Allcott & Genzkow, 2017; Meppelink et al., 2017; Howell et al., 2022). Confirmation bias (CB) has been named one of factors of users' susceptibility to untruthful information.

In psychology, CB has been mostly seen as a constant individual trait whose intensity may vary between individuals. Most broadly this trait be defined as the tendency to choose, interpret, search for, recall, and remember the information that supports agent's existing beliefs or expectations (see Cooper et al., 1970; Koriat et al., 1980; Wickens, et al., 2000; Peters, 2022).

In media research, CB is usually studied as a group-level effect. Its proven manifestations include, among other things, ascribing higher credibility (Moravec et al., 2018, Kim & Dennis, 2019), usefulness and convincingness (Meppelink et al., 2019) to attitude-consistent information.

While CB as the effect of message attitude-consistency on message believability has been widely investigated, the role of social cues, particularly user comments for it is still not quite clear, although their importance for news consumption, given the increasing presence of peer-to-peer communication functionalities on different platforms, has been widely acknowledged. Research shows that negative comments decrease trust in and perceived quality of news items (Waddell, 2018; Dohle et al, 2018) and that positive comments may have no effect of perceived news credibility (Kluck et al 2019). As the interaction of comment valence and news valence has not been studied, it is unknown under which conditions, if at all, comments may reinforce or reduce confirmation bias. Some research has even shown that only 40% of users read news comments at all (Steinfeld et al., 2016) which may question the detected and, especially, undetected effects of comments.

Moreover, it is still unclear whether confirmation bias actually affects the overall accuracy of discrimination between trues and fakes which, in terms of countering online misinformation, is the most important question. Since attitude-consistency may increase trust not only to fakes, but also to truths, and inconsistency, likewise, may decrease trust to both, fewer mistakes might be expected in users whose attitudes toward issues covered in messages are not too extreme, thus preventing them from uncritically accepting all attitude-consistent messages and rejecting all attitude-inconsistent ones. This effect, however, has not been tested yet.

Further, if CB affects accuracy of truthfulness detection, conditions under which CB increases should simultaneously be associated with decreased accuracy. Thus, news valence and issue attitude might have their independent effects on news believability, along with attitude consistency, thus introducing asymmetry of CB size across valence and attitude values. Revealing such asymmetries is a novel task per se and an additional way of testing the association of CB with the ability to detect fakes; still such studies are lacking.

We contribute to closing these gaps by an experimental study offering participants (N=50) to evaluate the veracity of news items each covering a socially divisive issue (abortion, death penalty and LGBTQ+). We embed news items in an interface imitating a

social media post with a re-posted news item and a user comment, thus manipulating four factors: message veracity (true / false), message valence (for / against the issue), comment valence (for / against /null) and the issue itself. Issue attitudes are self-reported; attention to comments is controlled via eye-tracking.

Thus, we not only investigate comment valence and its interaction with news valence, but also single out comments really noticed by participants, by employing an eye-tracking design that has never been used for this task before. Further, to account for possible CB asymmetries, we refrain from measuring attitude consistency as a single variable, as it was done before - either binary (Moravec et al., 2018) or scale (Knobloch-Westerwick et al, 2014; Sülflow et al., 2018; Kim et al., 2018; Figl et al., 2023) – and opt for measuring it as an interactive term of message valence and users' issue attitude. Finally, experimental approach combined with multilevel regression modeling allows us to calculate individual values of CB based on participants' actual experimental responses, rather than on their self-reported evaluations of their CB as in (Chaiken et al., 1980; Meppelink et al., 2019), and to directly determine their effect on news truthfulness detection accuracy.

2. Literature review and hypotheses

2.1. Confirmation bias

Different psychological approaches explain CB by different cognitive mechanisms, including predisposition of the human brain to minimize its cognitive effort by preferring quick intuitive decision making to that driven by elaborated logic. This preference may be formulated in terms of Kahneman's theory (Kahneman, 2011) of thinking System type 1 and System type 2 (see Moravec et al., 2018) or Chaiken's (Chaiken, 1980) heuristic-systematic model of information processing (HSM) (see e.g. Dunbar et al., 2014). Both theories link difficulties in decision making to cognitive load, which makes them related to the theory of cognitive dissonance, stating that people avoid information that makes them uncomfortable (Festinger, 1957). One of the sources of such discomfort might be mismatch between individual attitudes and message valence. Indeed, many studies have linked CB to the mechanism of cognitive dissonance (see e.g. Jonas et al., 2001; Knobloch-Westerwick et al., 2014; Moravec et al., 2018).

Thus, psychological theories suggest that confirmation bias effect is of adaptive nature (see Peters, 2022 for review). In contrast, media studies draw attention to its maladaptive aspects by showing that people tend to believe (Meppelink et al., 2019; Knobloch-Westerwick et al., 2014) and spread (Kim et al., 2019) messages consistent with their views even if they are labeled as false (Moravec et al., 2018;Gwebu et al 2022), while such labels effectively decrease sharing intention for attitude-inconsistent news (Gwebu et al 2022). Another maladaptive consequence of CB is formation of echo chambers – information environments where only belief-consistent messages circulate (e.g. Jiang et al., 2021) thus contributing to social polarization and conflict.

Attitude-consistent messages receive more reading time (Sülflow et al., 2018, Knobloch-Westerwick et al., 2014, Kim et al., 2019), higher scores of usefulness, convincingness, accuracy (Meppelink et al., 2019), and produce more willingness to like,

share and comment on them supportively (Kim et al., 2019). Most importantly, attitude-consistency shows positive relationship with perceived news credibility across multiple studies (Kim & Dennis, 2019, Kim, 2019, Figl et al., 2023, Knobloch-Westerwick et al, 2015; Westerwick, 2017 Kim et al., 2019, Moravec et al., 2018). Such messages may be rated as more credible even if they are labeled as false (Moravec et al., 2018). Therefore, our first and the most basic hypothesis we aim to test whether this well-known effect will reproduce:

H1. The higher the consistency of a news item with the reader's attitude to the covered issue, the higher is the perceived credibility of the news item.

H1a. The higher user's support of an issue, the higher is the perceived credibility of a news with pro-issue valence.

H1b. The higher user's support of an issue, the lower is the perceived credibility of a news with counter-issue valence.

It is important to note that news messages, unlike opinion pieces or comments, seldom carry explicit author attitudes towards covered issues; instead they may cover the events that are either desirable or undesirable for issue supporters (e.g. abortion ban and legalization have different desirability for pro-life news consumers). Therefore we define CB in relation to news as desirability bias (Sharot & Garett 2016; Tappin et al 2017), i.e. the inclination to believe in desirable events more than in undesirable.

2.2. Attitude-based and news-valence-based asymmetry of confirmation bias

Some research suggests that the strength of relationship between attitude consistency and message believability may be different in the different areas of the scale measuring readers' attitude. Thus, Kim et al (2019) find that news rating measured as a trinary variable has an effect of news believability only when it is low (as compared to medium), but not when it is high. Likewise, well-studied higher popularity of "bad news" as compared to good news, also known as negativity bias (Knobloch-Westerwick et al 2020), suggests that news on counter-issue events, such as legal bans and protests, might induce higher believability irrespective of readers' issue attitudes or that, at least, the effect of the latter might interact with news valence. However, despite proved user preference for negative news, the few existing studies have obtained no evidence of their higher believability (Coutts 2018; van der Meer & Brosius 2024). Since the relevant research is lacking, we formulate both of our assumptions as research questions:

RQ1: How, if at all, the relationship between issue attitude and perceived news credibility, will differ in different areas of issue attitude?

RQ2: How, if at all, the strength of the relationship between news attitude consistency and news believability will differ for different types of news valence?

2.3. Comment-valence-based asymmetry of confirmation bias

Social media user comments are an example of social (dis)approval signals that can serve as heuristic cues for attitude formation and update (Dohle et al., 2018). There is no consensus among researchers on the impact of comment valence on perceived credibility of the news. Kluek et al. (2019) find that negative comments questioning the credibility of news articles reduce perceived credibility ratings of news compared to a condition with no commentary, while positive comments produce no change in news believability, which is also consistent with earlier results (Waddell, 2018, Winter, Brückner, and Krämer et al., 2015). However, several other studies show that no type of comment valence, either positive or negative, is predictive of news credibility perceptions (Steinfeld et al., 2016, Dohle et al., 2018).

Such inconclusive results may have several explanations. First, studies differ by the absence or presence of control condition (no commentary) and by their definition of comment valence which is mostly understood as commentator's support of the news message, not his/her attitude towards an issue covered in it. More importantly, the studies reviewed do not track whether a comment has been read by the user at all. A few papers employing eye-tracking for this purpose show that the proportion of comment readers among participants is between 40% and 57.5% (Steinfield, 2016, Sülflow et al. 2019). These studies also find the effect of comment valence on perceived news credibility to be limited to one news issue out of three (Steinfeld et al 2016) and no effect on self-reported intention to click on the news (Sülflow et al. 2019).

No research investigates a possible interactive effect of comment valence and news valence on news believability, with or without eye-tracking, which leads us to the following hypothesis:

H2. The effect of the level of support of news valence on perceived news credibility is moderated by the level of support of the comment valence.

2.3. Confirmation bias and veracity detection

CB as the inclination to believe news based on their attitude consistency or desirability, instead of their actual veracity, can be expected to decrease human ability to discern true and fake news, unless non-biased individuals use even less effective detection strategies, such as random guesses. Therefore, our next hypothesis is:

H3. *The higher the individual confirmation bias, the lower the probability of correct news veracity detection.*

The potential negative effect of CB on accuracy of news truthfulness detection may also be traced indirectly: first, via the higher error probability under the conditions with increased CB (assumed in RQs 1 & 2) and, second, via the effects of related concepts. Thus, the review by Howe (2017) offers the following mechanism: belief extremity caused by subjective issue importance hinders human ability to change their beliefs. We then can expect that the stronger and less changeable beliefs will contribute to a larger gap in the levels of trust to belief-consistent and inconsistent messages, that is, to CB, and, consequently, to a weaker ability to evaluate the truthfulness of news. However, existing findings regarding this are

mixed. When extreme views are shown to correlate with susceptibility to fake news, these are usually extreme right views, as they are understood in the North-American and European cultures (see Baptista&Gradim (2022) for a review). In the Chilean context, Halpern et al (2019) find no relationship between political views strength and subjective credibility of fake news, while a study on Hong Kong respondents even finds that holders of certain extreme views (associated with anti-patriotism) are in fact better in identifying misinformation (Au et al 2021).

Given the scarcity and inconclusiveness of such research, we base our next hypothesis on our interpretation of Howe's theory outlined above:

H4. Attitude extremity will be negatively related to the news truthfulness detection accuracy

At the same time, the findings about asymmetric contribution of belief, or attitude extremity of different types, suggest that it might affect accuracy differently depending on the degree of attitude consistency with both news and comment valences. As neither this interaction effect nor its mechanism have been described so far, we formulate our assumption as our last RQ:

RQ3: How, if at all, the strength of relationship between attitude extremity andnews truthfulness detection accuracy, will vary across different levels of news and comment attitude consistency?

2.5. Other factors of news veracity detection

Of several widely studied factors of fake news detection, two – rational thinking and media literacy – are of special relevance for our research. Rational or analytical thinking is usually understood as the ability to suppress intuitive ('system 1') decision making and to activate logic and reason ('system 2') instead. Measured mostly with Cognitive Reflection Test (CRT) (Frederick, 2005) it is consistently found to be related to correct information evaluation. Thus, Bronstein et al (2019) show that it largely determines the ability to identify false news, while a tendency towards dogmatism reduces the quality of information assessment. Likewise, individuals who are more likely to suppress intuitive thinking with further reflection are found to be better in fake news detection by Pennycook & Rand (2019) and those with poorer cognitive reflection are shown to form stronger belief in false headlines by Calvillo et al. (2021).

Media literacy is a domain-specific 'companion' concept for rational thinking and, most broadly, aims to capture ability to critically assess news messages produced by professional media organizations. Approaches to its measurement are dramatically different, including experimental media literacy interventions, knowledge and skill tests, self-assessment scales and self-reported employment in media industry. While interventions consistently show a positive short-term effect on fake news recognition (see Lu et al 2024 for a meta-review), the results of studies using survey-based measurement vary as much as the measurements themselves. News media literacy scales developed by the team of Ashley, Craft, Marksl,

Tully and Vraga (e.g. Ashley et al 2013) and their variations have been usually shown positively related to false news discernment, both by the authors themselves (Ashley et al 2023) and in other works (Chan 2022). However, Jones-Jang et al (2019) find out that neither Ashley's news literacy, nor BohPodgornik's information literacy are related to false news recognition, while Inan&Temur's is. A comparison of media professionals with ordinary social media users reveals no difference in fake news recognition, unless both groups get involved in fact checking which increases the accuracy of professionals, but not of the lay people (AUTHOR, 2021). In this research, our general expectation is that both rational thinking and media literacy are likely to improve fake news recognition and therefore should be controlled for.

3. Method

A laboratory between-subjects experiment was conducted, with participants' eye movements being recorded while they were reading news messages. Participants were asked to evaluate news veracity using a 6-point Likert scale. Each news item was embedded in an interface resembling that of a typical social media platform (see Figure 1) and accompanied by a user commentary. The 2x2x3x3 experimental design involved manipulating the following variables: *news valence* (pro-issue vs counter-issue), *commentary valence* (pro-issue vs counter-issue vs. no comment), *news veracity* (true vs false) and *news issue* (LGBTQ+ rights, death penalty or abortion), the two latter having been used as controls. The study also included collection of socio-demographic information about the participants, Cognitive Reflection Test (CRT), media literacy assessment test and an interview.

The research design was approved by the Ethics Committee for Empirical Research Projects (HSE University).

Figure 1. Example of an item as presented to participants (translated from Russian): news issue – death penalty; veracity – true; news valence – counter-issue; commentary valence – counter-issue

3.2. Stimuli

We construct a collection of news (N=36, exclusive of one training item) with an even distribution of texts by veracity, valence and issue, thus consisting of 12 subsets each containing three items with identical parameters (e.g. the collection has three false news texts covering abortion with a pro-issue valence). Some of these texts were re-used from the Laboratory's earlier collection, and others added following exactly the same procedure.

True news were drawn from news outlets and cross-checked with other sources; fake news were created by research team members and edited by a professional journalist. The three issues (abortion, death penalty and LGBTQ+ rights) were selected from among the most socially divisive topics with a clear-cut split of public opinion[1]. Attitudes to these three issues are also conceptually important as they are widely used to differentiate between liberal and conservative, or (post)modern and traditional sets of values globally. Several issues were selected to decrease possible effect of each issue on the results. Only single-issue news items were included in the final database. Since news authors rarely demonstrate a pro- or counter-issue attitude explicitly, pro-issue valence was assigned to news items covering events desirable for issue supporters or undesirable for issue opponents (e.g. legalization of death penalty or abortion), while counter-issue valence label was used for news about the opposite types of events, such as bans on respective activities.

Each news item was matched to two unique comments (pro-issue or counter-issue, N = 72). To make comments look as natural matches for news items, comments for true news were selected from those that really had been posted under the respective news in the media outlets of their origin; for fake news, real comments were drawn from those accompanying similar real news in different media. The final database included 108 stimuli clustered in 36 triads with each triad containing one unique uncommented news text, the same text with the matched pro-issue comment and the same text with the matched counter-issue comment.

The average news text length is 334*(SD = 73)*. No differences in length between any of the main text groups (differing by news issue, veracity, news valence, and comment valence) are significant, as shown by Mann-Whitney and Kruskal-Wallis tests.

3.3. Participants & Recruitment

Russian native speakers (N=50, 17 male, mean age = 26,86 *(7,09)*) were involved as participants. Of them, 42% were studying, 46% were working, and the rest were unemployed

[1] According to the polls closest to the data collection time, abortion was unconditionally supported by 37% of Russians, while 13% were for unconditional ban (June 2022, https://wciom.ru/analytical-reviews/analiticheskii-obzor/preryvanie-beremennosti-za-protiv-i-kakova-rol-gosudarstva); 37% were for keeping the moratorium for death penalty or for the complete ban thereof, with 57% wishing to restore or even expand the use of death penalty (June 2021, https://www.levada.ru/2021/06/25/smertnaya-kazn-i-prestupnost/); 33% thought gays and lesbians should have the same rights as other citizens, but 59% held the opposite view (October 2021, https://www.levada.ru/2021/10/15/otnoshenie-rossiyan-k-lgbt-lyudyam/). The overall trend of the recent years has been conservative (towards support of death penalty and away from support of abortion and LGBTQ+ rights).

or retired. We excluded individuals employed in professions related to political science, sociology, psychology and media.

Participants were recruited via targeting through an advertising account on VKontakte social network as well as through invitations in social media groups, on universities' information boards and various information sites. After filling out a registration form, a participant was contacted by the experimenter and offered to choose a convenient time for the experiment.

3.4. Environment and equipment

Roughly a half (26) of participants were invited to the laboratory premises, while 23 individuals were subjected to the same experiment in a public space, usually a quiet and well-lit coffee shop. This was done to create a subsample treated under the condition of higher environmental validity: public spaces with a more relaxed and distracting atmosphere are much more common places for news consumption than research labs. Literature suggest that this may affect news perception: thus, Kim et al (2019) find out that people in hedonic mood, as opposed to utilitarian mood, are less critical and less likely to pay attention to news sources; therefore, experiment location was used as a theoretically grounded control variable. Irrespective of location, all participants were treated a cup of coffee after the experiment.

Web application for news presentation was created using several Python frameworks for the server, or backend, and JavaScript and HTML/CSS frameworks for the client side, or frontend. The frontend, mimicking un unclickable interface of social media network with one news item, contained a clickable Likert scale of perceived news veracity and QR codes for identifying areas of interest (AOI) in the video stream. The backend allowed random selection of news-comment combinations and random item order.

The recording of eye movements was done in a sequential setup with *Pupil Core* mobile eye-tracking glasses system (Pupil Labs, Pupil Labs GmbH, Germany). It was connected to the laptop with *Pupil Labs* v.3.4 open-source software (*Pupil Capture* for recording and *Pupil Player* for video viewing and preprocessing). The laptop was equipped with a 15,6-inch monitor (1900 x 1200) with sampling rate of 120 Hz.

Pupil Player built-in functions were used to define two main AOI, message area and commentary area (see Figure 2), and to compute gaze fixation events (using the Fixation detection plugin, minimum duration = 60 ms, maximum = 2000 ms.). An additional plugin was developed in Python for correcting fixation positions.). An additional Python plugin was developed to compensate for calibration decay in between re-calibrations performed after each four items. The plugin is available in the Social & Cognitive Informatics Lab repository on OSF (https://osf.io/m5dp6/?view_only=ecbe71d3ca0c47ff8c8d1ab7e73d695e).

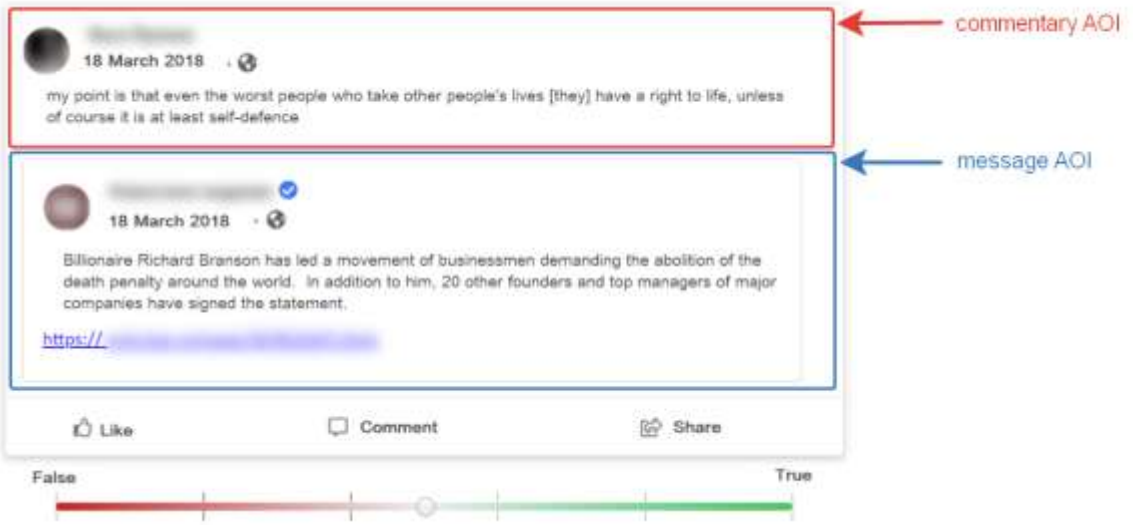

Figure 2. Example of an experimental screen with marked areas of interest

3.5. Experimental procedure

Participants signed an informed consent form before starting the experiment and received instructions before, after and in between the passes. In the beginning participants answered questions about their socio-demographic status and their attitudes towards abortion, LGBTQ+ rights and death penalty.

The eye tracking session started with the training text (not used in the analysis), after which each participant was shown all 36 news texts in random order. Each news text was randomly drawn from the respective triad, either with a pro-issue comment, or with a counter-issue comment, or without a comment, so that each of these three commentary types would be equally represented in the set seen by each participant.

A screen with the reference point appeared prior to the screen with each new stimulus. In line with the recommendations of Ehinger et al. (2019), eye tracker recalibration was carried out every 6 texts (~3-4 minutes). Given the size of our AOI (see Figure 2 below), the accuracy of < 0.5 and the precision of < 0.2 were considered sufficient. The eye tracking session had no time limit (min = 9:32 min, max = 14:45 min). After it, participants took a media literacy test and the Cognitive Reflection Test.

Finally, during the cognitive interview, the experimenter checked how well the participants understood the task and asked for their feedback regarding its difficulty. They were then encouraged to reflect on their strategies of distinguishing between trustworthy and untrustworthy news, and on the role of various news elements, comments and other details for their decisions. The session ended with a debriefing disclosing the veracity status of all news items to participants if they were interested.

3.6. Variables and measurements

Dependent Variables

Perceived message credibility rating (credibility): was measured as a score assigned by participants to news items on a 6-point Likert scale, from 1 (*definitely false*) to 6 (*definitely true*), with middle values denoting different degrees of participants' confidence in their responses.

Fake new detection accuracy (accuracy (binary)) measured the correctness of evaluation of veracity of each news item by participant. It was constructed by comparing binarized*credibility* values (where *false* = 1, 2 & 3; *true* = 4, 5, & 6) with *veracity* values as follows: if *credibility = veracity*, *accuracy binary = correct*, otherwise – *incorrect*.

*Fake new detection accuracy (accuracy (scale))*measured the degree of correctness of news item veracity evaluation by participant. It was constructed by transforming the 6-point *credibility* scale into the 6-point *accuracy (scale)* where 1 came to denote the least correct response, and 6 came to mean the most correct response, as was determined through comparison with news *veracity* values.

Independent and control variables

Veracity (control): true and false.

News issue (control): abortion, death penalty (DP) and LGBTQ+.

News valence: pro-issue and counter-issue. Pro-issue news are those covering events desirable for issue supporters and undesirable for issue opponents; counter-issue news are the opposite.

Commentary valence: pro-issue, counter-issue and no comment (nocomm). Pro-issue comment directly supports the issue or the pro-issue event or opposes the counter-issue event; counter-issue comment opposes the issue or the pro-issue event or supports the counter-issue event.

Commentary presence: yes and no.

*Attitude toward issue (attitude):*respondents were asked to express their attitudes towards abortion, LGBTQ+ and the death penalty on a 7-point Likert scale ranging from 1 (*definitely oppose)* to 7 (*definitely support).*

Strength of the users issue attitude (attitude extremity) was constructed by transforming *attitude* scale into a 4-point scale ranging from 0 obtained by recoding the middle *attitude* value (4) to 3 obtained by recoding any of the extreme *attitude* values (1 or 7).

News attitude consistency (in accuracy prediction models and in model 3): the degree of similarity between news valence and issue attitude (6 – maximal consistency, i.e. attitude = 7 and valence = pro-issue or attitude = 1 and valence = counter-issue; 0 – minimal consistency, i.e. the opposite combinations)

Comment attitude consistency (in accuracy prediction models): the degree of similarity between comment valence and issue attitude; measured as the previous.

News and comment attitude consistency (in credibility prediction models): evaluated via interaction between attitude, news valence and comment valence.

Individual confirmation bias (ICB): evaluated with regression modeling (see Data Analysis).

Rational thinking (CRT)(control)was measured with Cognitive Reflection Test (Frederick, 2005) which contains three brainteaser items using a free-response format. CRT translated and adapted versions were evaluated by experts and pre-tested (N=100, M_{age} = 26,9 *(10,4)*, Cronbach's alpha = 0.82). Further analysis uses participants' average values of CRT .

Media literacy (control) was measured with news media literacy scale (Ashley et al., 2013 translated into Russian and adapted for a Russian-speaking audience. In accordance with the original methodology, the final scale contained 15 questions grouped into 4 domains. These questions aim at testing participants' understanding of news production, dissemination and perception mechanisms, without involving respondents in self-assessment of their skills and avoiding country-specific issues, which is why this scale became the tool of our choice. Like CRT, it was evaluated by experts and pre-tested (N= 100, M_{age} = 26,9 *(10,4)*, Cronbach's alpha = 0.96). Average values were taken into the further analysis.

Trial duration (control) is the duration of one a trial from the moment the screen appears until the respondent presses *next* button.

Duration(control) is the ratio between the time a participant fixates their gaze on an OAI within a trial and the overall *trial duration*. This metric allows reliably comparing time spent on texts of different lengths as well as on texts presented with and without commentary.

Place (control): in lab and outside lab

3.7 Data analysis

Statistical analysis was carried out with lme4 package (Bates et al, 2015) in the R environment (R Core Team, 2022). Commented data analysis code and the dataset itself can be found on the OSF repository (https://osf.io/m5dp6/?view_only=ecbe71d3ca0c47ff8c8d1ab7e73d695e). Statistical significance was assessed using linear mixed models[2] (LMM) with individual (*n*=50) and item (*n*=36) specified as crossed random factors contributing 1800 observations. Model diagnostics showed interclass correlation coefficients above the conventional threshold for including random effects. The models also passed the tests for autocorrelation and heteroscedasticity; as the normality requirement was not entirely met, we can expect our estimates to be less precise, still unbiased (Schielzeth et al 2020).

Simr package (Green & MacLeod, 2016) was used for model power analysis. It showed that with our sample size the probability of detecting the effects of the size estimated by our models is 99.7% in *credibility* prediction, and 90% in *accuracy (scale)* prediction. In both groups of models, the best model was defined based on the Akaike information criterion (AIC) in the course of bottom-up stepwise selection. Fixed-effect estimates AICs and BICs for all models are presented in Appendix 1.

The scheme of the bottom-up stepwise selection between credibility (left) and accuracy (right) prediction models is shown in Figure 3. During preliminary analysis, it was found out that it is not comment valence, but the mere presence of any comment that is important for

[2]In the paper we present LMM results, but the code for Cumulative Link Mixed Models (CLMM), widely used for this type of tasks along with LMM, is also available in the mentioned Laboratory's GitHub. Effect sizes, directions and significances obtained with CLMM are broadly the same.

predicting news credibility. We also found out that the condition with negative comment was significantly different from the condition without a comment (b = 0.442, t = 2.575), but the difference between the conditions with positive and negative comments was not statistically significant (b = 0.250, t = 1.402). As it can be seen from Figure 3, the model predicting credibility with *commentary presence* instead of *commentary valence* demonstrates the highest quality; it has been selected as final.

Attitude consistency in this group of models was entered not as a separate variable, but (initially) as two interaction terms: *issue attitude * news valence* and *issue attitude * comment valence*, both allowing to differentiate between consistent and non-consistent conditions (need for H1), as well as between conditions where attitudes are (in)consistent with pro- and counter-issue news and comment valences (needed for RQ2 and H3). Confirmation bias was evaluated as a group level phenomenon via significance of the relationship between each of these two interaction terms and the outcome variable (*credibility*). In the final model interactions included commentary presence.

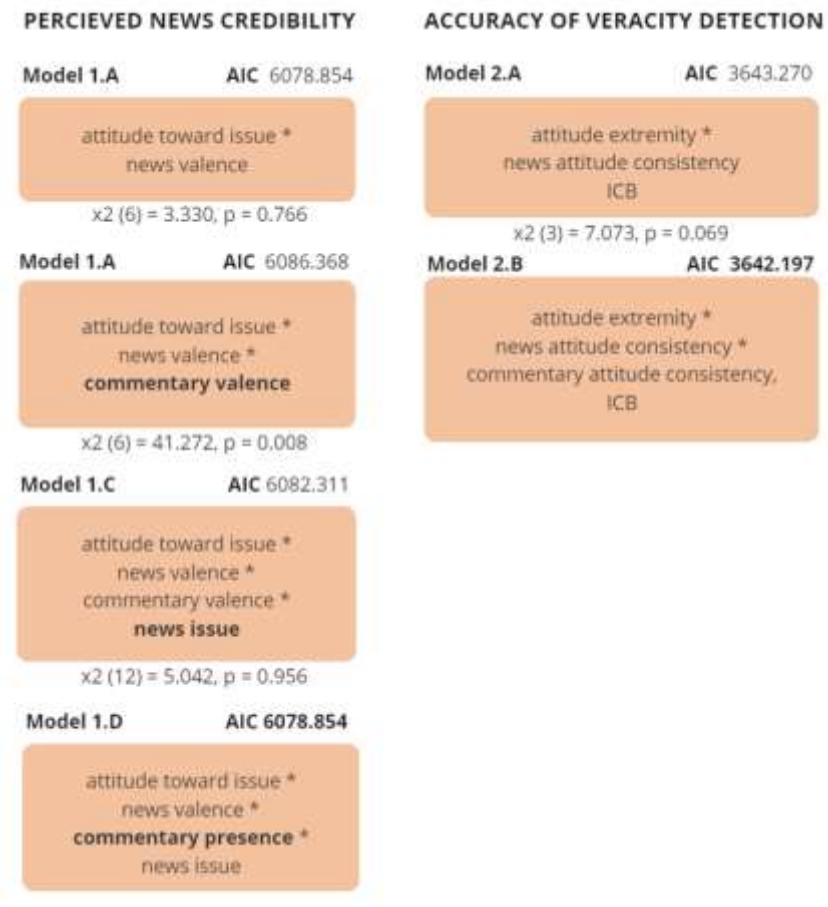

Figure 3. Stepwise model selection scheme

As the power analysis showed that we do not have enough power to analyze *accuracy (binary)*, only *accuracy (scale)* models were further tested and interpreted. Selection scheme was similar to that of credibility models; however, in accordance with our hypotheses, the main predictors were *attitude extremity*, news and comment *attitude consistency*, their interactions and *ICB. ICB* values were modelled as coefficients of individual-level random slopes in a separate LMM predicting *credibility* with *news attitude consistenc*y as an explicit variable (not as an interaction term) (see Table 7 with Model 3 in Appendix 1).

4. Results

4.1. Overall descriptive statistics and unnoticed comments

A total of 1800 observations were obtained from our sample of 50 subjects. Table 1 shows that, in line with our previous research, our participants are slightly better than random in differentiating between true and false news. This divergence from random guess is significant, as is shown in regression where news veracity reliably predicts perceived truthfulness (Table 2). Participants also hold quite strong views on the three studied issues.

Table 1. Main individual-level measurements

	Mean	SD	Min	Max
Age, years	25.469	6.533	17	45
Media literacy (range: 1-7)	5.866	0.544	4.333	6.733
CRT (range: 0-1)	0.591	0.397	0	1
Avg. accuracy (scale) (range: 1-6))	3.758	0.291	3.166	4.472
Avg. attitude strength (range: 0-3)	2.165	0.593	1	3
Avg. trial duration, ms.	720.855	134.908	473.741	1347.878

Table 2. Main news item-level measurements grouped by news issue

Topic	Attitude (range: 1-7)	Attitude strength (range: 0-3)	Credibility (range: 1-6)	Accuracy (range: 1-6)	Processing time, ms
Abortion	5.714 *(1.668)*	2.259 *(0.818)*	3.935 *(1.572)*	3.709 *(1.613)*	710.236 *(186.318)*
LGBTQ+	5.612 *(1.760)*	2.241 *(0.866)*	3.645 *(1.509)*	3.616 *(1.51)*	716.041 *(188.426)*
Death penalty	2.469 *(1.656)*	2.037 *(1.002)*	3.730 *(1.552)*	3.946*(1.507)*	711.376 *(180.634)*
All topics	4.598 *(2.282)*	2.178 *(0.904)*	3.768 *(1.547)*	3.757 *(1.549)*	712.652 *(185.137)*

Note: Cells contain means and standard deviations

As shown in table 2, our sample is also likely to be somewhat less conservative than the general Russian population, with lower support for death penalty and higher support for the two other issues. Additionally, although there are few significant correlations between independent variables, pro-LGBTQ+ attitude is positively and strongly related to pro-abortion attitude (r=0.62), and negatively – to support for death penalty (r=-0.358) and to respondent

age (r=-0.327). There are no significant differences between news on different issues either in average accuracy of false message detection or in the mean time spent by users.

The proportion of trials with unnoticed comments, as determined with the eye-tracking technology, was less than 10% of all trials with comments. These trials were coded and analyzed as *no commentary* condition, since users could not be affected by them.

4.3. Hypothesis testing. Message credibility

Table 3 lists the fixed-effect terms of the model predicting perceived message credibility. Non-significant effects for 3 & 4-way interaction terms are reported in the Appendix 1 only. Control variables (listed first) show that females and older participants have higher levels of trust in news messages, while media literacy and rational thinking are predictably unrelated to it.

Table 3. Fixed-effect estimates with t-values of the Linear Mixed-Effect Model predicting perceived credibility of news

	Perceived credibility		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.03	2.58 – 3.48	<0.001
Age	0.17	0.10 – 0.25	<0.001
Gender [Male]	-0.18	-0.35 – -0.01	0.034
Media Literacy	-0.01	-0.08 – 0.07	0.871
CRT	0.05	-0.02 – 0.12	0.194
Attitude	-0.08	-0.42 – 0.27	0.665
Veracity [true]	0.46	0.24 – 0.69	<0.001
News valence [pro-issue]	0.97	0.37 – 1.58	0.002
Commentary presence [yes]	0.10	-0.41 – 0.61	0.693
News issue [LGBTQ+]	0.28	-0.33 – 0.88	0.373
News issue [DP]	0.77	0.08 – 1.46	0.030
Attitude × News valence [pro-issue]	0.64	0.15 – 1.13	0.010
Attitude × Commentary presence [yes]	0.24	-0.21 – 0.69	0.295
Attitude × News valence [positive] × Commentary presence [yes]	-0.73	-1.37 – -0.09	0.025
Attitude × News valence [pro-issue] × Commentary presence [yes] × News issue [LGBTQ+]	1.04	0.15 – 1.92	0.022

Attitude × News valence [pro-issue] × Commentary presence [yes] × News issue [DP]	1.75	0.83 – 2.67	<0.001
Conditional R^2 / Marginal R^2	0.200 / 0.101		

Note: Bold - significant effects

As expected, we find no main effect of issue attitude; however, pro-issue news turn out to be trusted more than counter-issue news, irrespective of users' issue attitude on average (b = 1.009, t = 3.365, p = 0.002). Most importantly, the coefficient of the interaction term involving these two values (which measures attitude consistency) is significant which means that the increase in issue support in users leads to the increase in credibility of pro-issue messages and to the decrease in credibility of counter-issue items (see Figure 4). This confirms H1a, b about the existence of confirmation bias.

Figure 4 also shows that the difference in perceived credibility of pro-issue and counter-issue news is significant only among issue supporters, but not among those opposing respective issues. In other words, the importance of attitude consistency for news believability increases with issue support. This provides evidence in favor of attitude-based asymmetry of CB and a response to our RQ1.

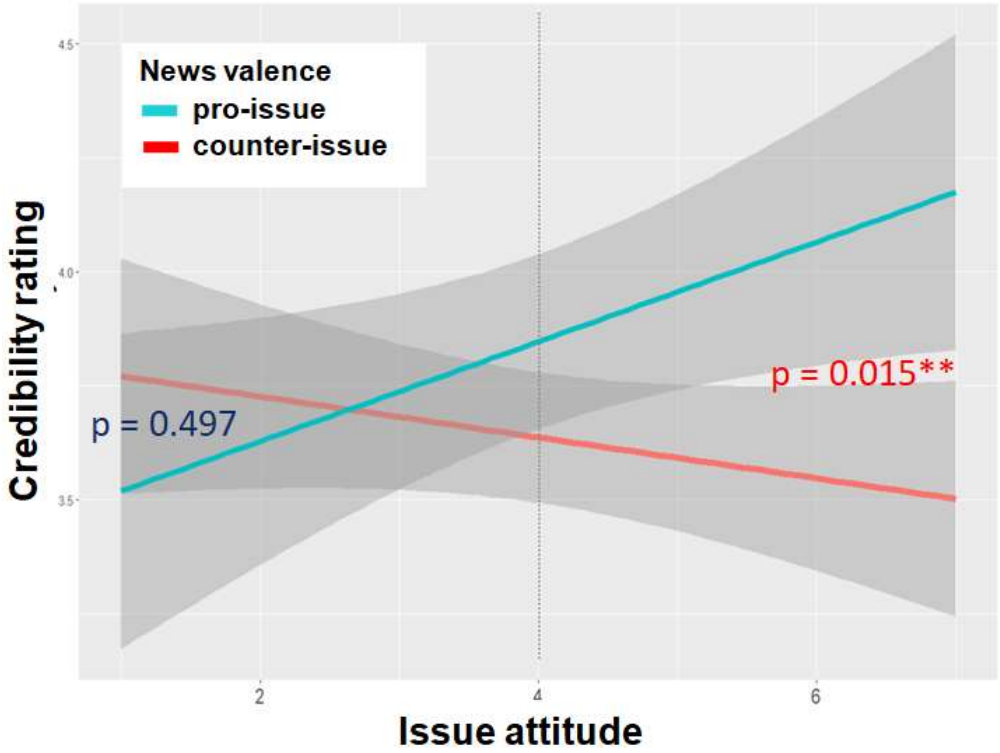

Figure 4. Confirmation bias with attitude-based asymmetry: relationship between perceived message credibility and issue attitude

Our next question (RQ2) was whether importance of attitude consistency for believability would vary depending on news valence. Figure 5 shows that perceived credibility of both pro-issue and counter-issue news grows with the growth of attitude consistency. Although this growth looks smaller for counter-issue news, there is no sufficient evidence that this difference is significant, which speaks against news-valence-based

asymmetry in confirmation bias. At the same time, as shown by post hoc comparison analysis (contrasts), consistency-motivated increase in believability of counter-issue news (red line) is not statistically significant either (p = 0.206), while that of pro-issue news (green line) is (p = 0.082). This evidence speaks for the opposite conclusion, in favor of asymmetry: users tend to obtain biased visions of news truthfulness only if their views are consistent with the valence of pro-issue news, but not with that of counter-issue news.

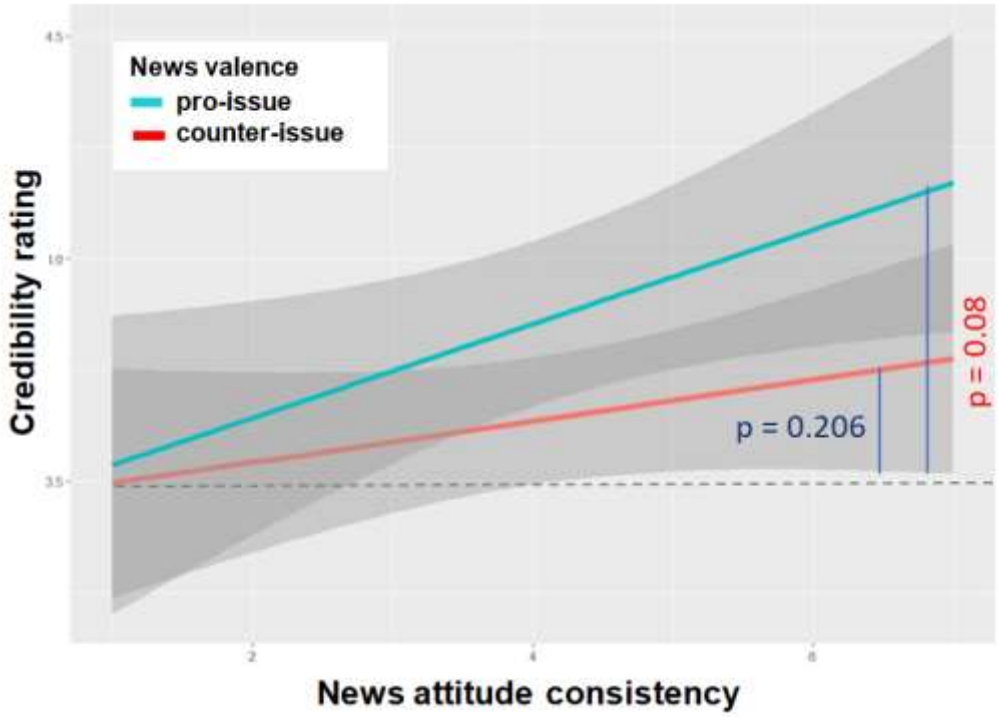

Figure 5. Relationship between attitude consistency and perceived news credibility

As mentioned above, we did not find either the main effect of commentary valence on perceived news credibility or the effect of the interaction between commentary valence and issue attitude thereon (see models 1.A-C in appendixA). This means that we have no evidence of commentary valence participating in CB formation. Therefore, comment attitude consistency cannot moderate the effect of news attitude consistency on news credibility which leads us to reject hypothesis 2.

At the same time, statistically significant difference between conditions with counter-issue comments and with no comment has given us a reason to explore the role of commentary presence instead of commentary valence. Specifically, we have assumed that issue attitude might affect perceived news credibility differently not only depending on news valence, but also depending on the presence of a comment, irrespective of the valence of this comment.

As it is seen from Table 3, the effect of the three-way interaction between issue attitude, news valence and comment presence is significant, thus providing proof for our assumption. Figure 6 shows the impact of the commentary presence on the perceived credibility in more detail. It offers suggestive evidence that in no comment condition news valence affects the trust of users with strong pro-issue attitudes (which is in line with our findings regarding

RQ1), while the presence of comments makes users with modest pro-issue attitudes sensitive ϕto news valence.

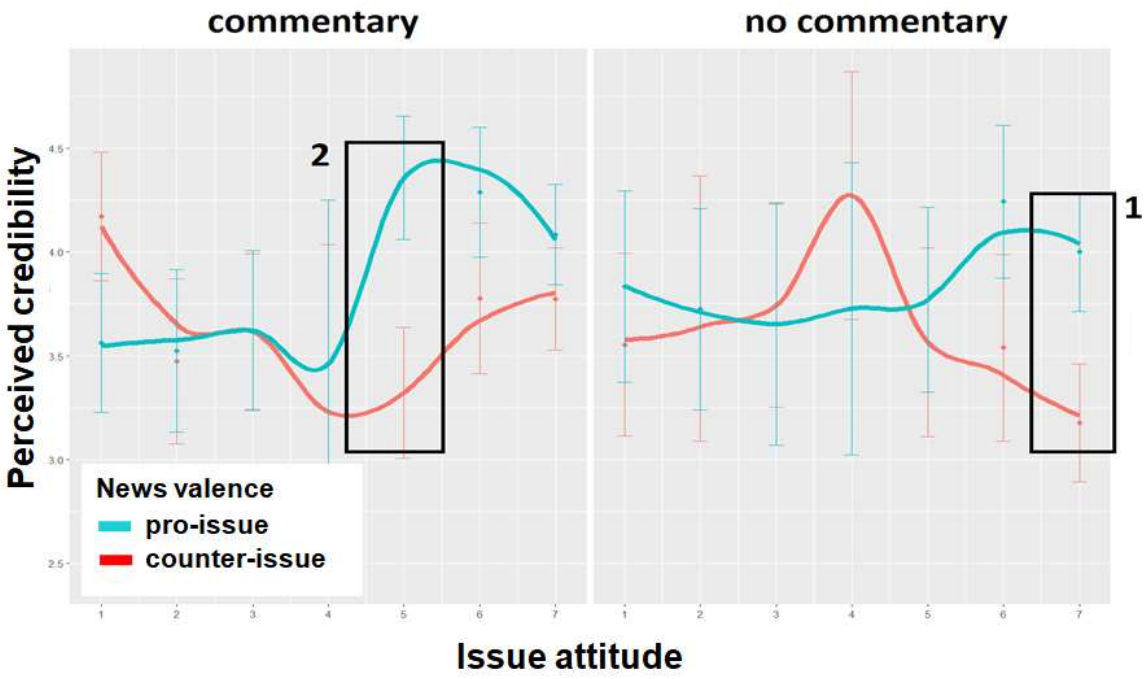

Figure 6. Dependence of perceived message credibility on issue attitude, news valence and commentary presence

4.4. Hypothesis testing. Fake news detection accuracy

Table 4 lists the fixed-effect terms of the model for message veracity detection accuracy (scale). Non-significant effects for 2&3-way interaction terms are reported in the Appendix 1 only. Out of all control variables, only veracity has a significant effect on accuracy: participants are slightly better at detecting the truth status of true news than that of false news. This might happen because, unlike fabricated news, true news may have been encountered by users before the experiment. Also, absence of the effect of age and gender on accuracy, combined with the presence thereof on perceived credibility, tells us that lower overall trust does not give younger men any advantage or disadvantage in fake news detection ability; therefore, it is likely to substitute type 1 errors with type 2 errors.

.

Table 4. Fixed-effect estimates with t-values of the Linear Mixed-Effect Model predicting scaled accuracy of news truth status detection

	Accuracy (scale)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	5.02	2.51 – 7.53	<0.001
age	-0.01	-0.11 – 0.08	0.788
gender [Male]	-0.06	-0.27 – 0.15	0.553
ML	0.03	-0.06 – 0.12	0.506
CRT	-0.03	-0.11 – 0.06	0.566
Veracity [true]	0.64	0.35 – 0.93	<0.001
attitude extremity	-0.56	-1.39 – 0.28	0.190
News attitude consistency	-0.28	-0.87 – 0.31	0.359
Comment attitude consistency	-0.55	-1.14 – 0.04	0.067
news issue [lgbt]	0.03	-0.33 – 0.39	0.863
news issue [penalty]	0.30	-0.06 – 0.66	0.101
ICB	-0.36	-0.44 – -0.27	<0.001
attitude extremity × comment attitude consistency	0.20	0.00 – 0.39	0.045
Marginal R^2 / Conditional R^2	0.118 / 0.232		

Note: Bold - significant effects

We can see that individual confirmation bias is significantly and negatively related to accuracy, which means that less biased news consumers are better in discerning true and fake news. We have full support for H3.

Attitude extremity is related to accuracy in the expected direction (negatively, thus seemingly subverting human ability to discern true and false news), however, this relationship is not significant, and H4 cannot be confirmed. Neither we find any meaningful interactions between attitude extremity and attitude consistency. The only significant effect suggests that individuals with weak attitudes tend to be better in discriminating between true and false news accompanied by counter-attitude comments than those accompanied by pro-attitude comments, but no such distinction if found for individuals with moderate or strong attitudes. We have to conclude that neither attitude extremity nor attitude consistency affect news truthfulness detection accuracy either independently or in a combination, which is a response to our RQ4.

Finally, additional regression modeling (not shown) has found no relationship between attitude and news truthfulness detection accuracy. This speaks against our assumption about possible decrease of accuracy under conditions that increase CB (in this case – higher pro-issue attitude). However, this negative result seems to be of less importance than the detected main effect of ICB.

5. Discussion and conclusions

Using a laboratory experiment with eye movement recording, this study has investigated the relationship of confirmation bias to the ability to discern true and false messages and the role of social cues, namely comments, for CB reinforcement or compensation.

In contrast to earlier experimental results (Steinfeld et al., 2016; Sülflow et al. 2019) we find that the proportion of trials in which users paid attention to comments is very high (over 90%). This might result from the design of our experimental interface in which the comment preceded the news message.

While a separate study might investigate the effect of comment location on the strength of its influence on users, it would be logical to assume that comments preceding the message have a higher chance to affect news consumers than those following it. However, while we do find the effect of message valence on news perceived credibility (i.e. classical CB), we do not find any effect of comment valence thereon. A possible explanation for that might be related to CB's socially undesirable and, consequently, unintentional character. This character is likely to lead individuals to attempt eliminating the influence of CB on their judgements. Such self-control might be expected to be easier when message valence is formulated as an explicit issue attitude in a separate text (comment) and harder when it is represented in the form of (un)desirable event description within a (seemingly) objective news text. This may result in weaker or null effect of comment valence on news believability, as compared to the effect of news valence. Indeed, this conclusion is consistent with some of the participants' responses in their cognitive interviews, who noted that they "deliberately ignored the comments" because they felt that the comments were meant to interfere with their decision-making.

We, however, do find the effect of commentary presence or, rather, commentary absence on believability and thus on the character of CB: in the no-comment condition CB is more similar to the main CB effect in shape and more interpretable. The main CB effect demonstrates attitude-based asymmetry: specifically, it is significant only for users with pro-issue attitudes. It means that issue supporters, on average, have more difficulties in controlling CB than those who are against the issue. A possible psychological mechanism behind this might be that issues supported by an individual produce stronger emotions and, as a result, stronger wishful thinking than issues opposed by them. Thus, an abortion supporter, on average, will wish a news about abortion ban lifted to be true more strongly than an abortion opponent will wish the same news to be false. This mechanism, however, needs to be tested. Likewise, the shift of CB in comment-present condition from stronger issue supporters to weaker issue supporters suggests that comments add complexity to the process of news perception that requires further study to be fully interpreted.

Our results on modeling the accuracy of news veracity detection unequivocally tell us that it is subverted the more the higher the individual CB, i.e. that confirmation bias is maladaptive. Although it may help quickly resolve cognitive dissonance and relieve an individual from cognitive load, it simultaneously leads to errors in judgements about the veracity of certain events and thus may further result in mal-informed and disfunctional

decisions. A reservation should be made here that this conclusion is valid only for the CB type studied in this work - the inclination to believe in desirable events more than in undesirable – known as desirability bias (Sharot & Garett 2016; Tappin et al 2017). It might not apply to an inclination to believe in events or phenomena consistent with attitudes or prior domain knowledge (e.g. "A patient survived a fever of 50℃").

Simultaneously, it is perplexing that attitude extremity, while it directly defines the strength of confirmation bias, is not related to the accuracy of news veracity detection. One of the directions of further research here (apart from experimenting with longer extremity scales) might be investigating issues that have much higher personal importance, such as beliefs about diseases which individuals themselves suffer from. Following Howe's mechanism (2017) of belief formation, more important issues should evoke stronger beliefs and be more subversive for the ability to update them, which is why such beliefs may produce stronger CB and have a more pronounced negative effect on the accuracy of news veracity detection.

6. Limitations

Despite the sample of 1800 observations, 50 individuals may be insufficient to detect many individual-level effects and some instance-level effects if they are small. The sample is to some extent biased toward female, younger, more educated and less conservative participants, as compared to the general Russian population. The perception of comments in an experimental setting is probably not entirely ecologically valid as participants might be deliberately looking for the elements of the experiment designed to subvert their accuracy. Participants' domain expertise was not controlled for.

Funding details
Research was supported by the Basic Research Program at the National Research University Higher School of Economics (HSE University).

Declaration of competing interest
The authors report there are no competing interests to declare.

Data availability statement
The data that support the findings of this study are openly available in the Social and Cognitive Informatics Lab repository at OSF at (https://osf.io/m5dp6/?view_only=ecbe71d3ca0c47ff8c8d1ab7e73d695e).

References

Allcott, H., Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.

Ashley, S., Maksl, A., & Craft, S. (2013). Developing a news media literacy scale. *Journalism & Mass Communication Educator*, *68*(1), 7–21. https://doi.org/10.1177/1077695812469802

Ashley, S., Craft, S., Maksl, A., Tully, M., &Vraga, E. K. (2023). Can news literacy help reduce belief in COVID misinformation? *Mass Communication & Society*, 26(4). https://doi.org/10.1080/15205436.2022.2137040

Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021). Does political extremity harm the ability to identify online information validity? Testing the impact of polarisation through online experiments. *Government Information Quarterly*, 38(4), 101602.https://doi.org/10.1080/15205436.2022.2137040

Aufderheide, P. (2018). Media literacy: From a report of the National Leadership Conference on Media Literacy. In *Routledge eBooks*, 79-86. https://doi.org/10.4324/9781351292924-4

Baptista, J. P., &Gradim, A. (2022). Who believes in fake news? Identification of political (A)symmetries. *Soc. Sci.*, 11(10), 460. https://doi.org/10.3390/socsci11100460

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition*, *8*(1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005

Butler, L. H., Fay, N., & Ecker, U. K. H. (2023). Social endorsement influences the continued belief in corrected misinformation. *Journal of Applied Research in Memory and Cognition*, *12*(3), 364–375. https://doi.org/10.1037/mac0000080

Calvillo, D. P., Rutchick, A. M., & Garcia, R. J. B. (2021). Individual Differences in Belief in Fake News about Election Fraud after the 2020 U.S. Election. *Behavioral Sciences*, *11*(12), 175. https://doi.org/10.3390/bs11120175

Chaiken, S. (1980). Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion. Journal of Personality and Social Psychology, 39(5), 752-766.

Chan, M. (2022). News literacy, fake news recognition, and authentication behaviors after exposure to fake news on social media. *New Media & Society*. https://doi.org/10.1177/14614448221127675

Chauhan, K., & Pillai, A. (2013). Role of content strategy in social media brand communities: A case of higher education institutes in India. *Journal of Product & Brand Management*, 22(1), 40-51. https://doi.org/10.1108/10610421311298687

Coutts, A. (2018). Good news and bad news are still news: experimental evidence on belief updating. Experimental Economics, 22(2), 369–395. https://doi.org/10.1007/s10683-018-9572-5

Dohle, M. (2017). Recipients' assessment of journalistic quality. *Digital Journalism*, *6*(5), 563–582. https://doi.org/10.1080/21670811.2017.1388748

Dunbar, N. E., Miller, C. H., Adame, B. J., Elizondo, J., Wilson, S., Lane, B. L., Kauffman, A. A., Bessarabova, E., Jensen, M. L., Straub, S. K., Lee, Y., Burgoon, J. K., Valacich, J. J., Jenkins, J. L., & Zhang, J. (2014). Implicit and explicit training in the mitigation of cognitive bias

through the use of a serious game. *Computers in Human Behavior, 37*, 307–318. https://doi.org/10.1016/j.chb.2014.04.053

Ehinger, B. V., Groß, K., Ibs, I., & König, P. (2019). A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000. *PeerJ 7:e7086*. https://doi.org/10.7717/peerj.7086

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114–138. https://doi.org/10.1257/mic.3.2.114

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press. https://doi.org/10.1515/9781503620766

Figl, K., Kießling, S., & Remus, U. (2023). Do symbol and device matter? The effects of symbol choice of fake news flags and device on human interaction with fake news on social media platforms. *Computers in Human Behavior, 144*, 107704. https://doi.org/10.1016/j.chb.2023.107704

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Gupta, M., Dennehy, D., Parra, C. M., Mäntymäki, M., &Dwivedi, Y. K. (2023). Fake news believability: The effects of political beliefs and espoused cultural values. Information & Management, 60(2), 103745. https://doi.org/10.1016/j.im.2022.103745

Gwebu, K.L., Wang, J., &Zifla, E. (2022) Can warnings curb thespread of fake news? The interplay between warning, trust and confirmation bias.*Behaviour &Information Technology*, 41(16), 3552-3573, DOI: 10.1080/0144929X.2021.2002932

Halpern, D., Valenzuela, S., Katz, J., & Miranda, J. P. (2019). From belief in conspiracy theories to trust in others: Which factors influence exposure, believing and sharing fake news. In *Social Computing and Social Media. Design, Human Behavior and Analytics: 11th International Conference, SCSM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part I 21* (pp. 217-232). Springer International Publishing.https://doi.org/10.1007/978-3-030-21902-4_16

Howe, L. C., &Krosnick, J. A. (2017). Attitude strength. *Annual review of psychology*, *68*(1), 327-351.https://doi.org/10.1146/annurev-psych-122414-033600

Jiang, B., Karami, M., Lu, C., Black, T., & Liu, H. (2021). Mechanisms and attributes of echo chambers in social media.*arXiv(Cornell University)*.https://doi.org/10.48550/arxiv.2106.05401

Jiménez, Á. V., Mesoudi, A., & Tehrani, J. J. (2020). No evidence that omission and confirmation biases affect the perception and recall of vaccine-related information. *PLOS ONE*, *15*(3). https://doi.org/10.1371/journal.pone.0228898

Johnson, T. J., Kaye, B. K. (2015). Reasons to Believe: Influence of Credibility on Motivations for Using Social Networks. *Computers in human behavior,* 50, 544-555.https://doi.org/10.1016/j.chb.2015.04.002

Jonas, E., Schulz-Hardt, S., Frey, D., Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *J. Pers. Soc. Psychol*. 80(4), 557–571.https://doi.org/10.1037/0022-3514.80.4.557

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

Kahneman, D. (2011). Thinking, fast and slow. *Anchor Canada*.

Kim, A., & Dennis, A. R. (2018). Says who?: How news presentation format influences perceived believability and the engagement level of social media users. *Proceedings of the. Annual Hawaii International Conference on System Sciences*. https://doi.org/10.24251/hicss.2018.497

Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings. *Journal of Management Information Systems*, *36*(3), 931–968. https://doi.org/10.1080/07421222.2019.1628921

Kim, B., Xiong, A., Lee, D., & Han, K. (2021). A systematic review on fake news research through the lens of news creation and consumption: Research efforts, challenges, and future directions. PLOS ONE, 16(12). https://doi.org/10.1371/journal.pone.0260080

Klayman, J. (1995). Varieties of confirmation bias. The psychology of learning and motivation, 32, 385-418.

Kluck, J. P., Schaewitz, L., &Krämer, N. C. (2019). Doubters are more convincing than advocates. The impact of user comments and ratings on credibility perceptions of false news stories on social media. Studies in Communication, Media, 8(4), 446–470. https://doi.org/10.5771/2192-4007-2019-4-446

Knobloch-Westerwick, S., Johnson, B. K., &Westerwick, A. (2014). 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., Mothes, C., &Polavin, N. (2020). Confirmation Bias, Ingroup Bias, and Negativity Bias in Selective Exposure to Political Information. Communication Research, 47(1),104-124. https://doi.org/10.1177/0093650217719596

Koriat, A., Lichtenstein, S., Fischhoff, B. (1980). Reasons for Confidence. Journal of Experimental Psychology: Human learning and memory, 6(2), 107.

Ladeira, W. J., Dalmoro, M., De Oliveira Santini, F., &Jardim, W. C. (2021). Visual cognition of fake news: the effects of consumer brand engagement. Journal of Marketing Communications, 28(6), 681–701.https://doi.org/10.1080/13527266.2021.1934083

Liu, J. H. (1998). The Catastrophic Link Between the Importance and Extremity of Political Attitudes. *Political Behavior, 20* (2), 105-126. https://doi.org/10.1023/A:1024828729174

Lu, Chang; hu, bo; Bao, Meng-Meng; Wang, Chi; Bi, Chao; Ju, Zing-Da. Can Media Literacy Improve Fake News Discernment? A Meta-analysis. Cyberpsychology, Behavior, and Social Networking, Vol. 27, No. 4 (2024)

Lu, C., Bao, M. M., Wang, C., Bi, C., & Ju, X. D. (2023). Can Media Literacy Interventions Improve Fake News Discernment? A Meta-Analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4377372

Marquart, F., Matthes, J., & Rapp, E. (2016). Selective exposure in the context of political advertising: A behavioral approach using eye-tracking methodology. International Journal of Communication, 10, 20.

Mendel, R., Traut-Mattausch, E., Jonas, E., Leucht, S., Kane, J. M., Maino, K., Kissling, W., &Hamann, J. (2011). Confirmation bias: why psychiatrists stick to wrong preliminary

diagnoses. Psychological Medicine, 41(12), 2651–2659. https://doi.org/10.1017/s0033291711000808

Modgil, S., Singh, R., Gupta, S., &Dennehy, D. (2021). A confirmation bias view on social media induced polarisation during Covid-19. Information Systems Frontiers. https://doi.org/10.1007/s10796-021-10222-9

Moravec, P., Minas, R., & Dennis, A. R. (2018). Fake news on social media: People believe what they want to believe when it makes no sense at all. Kelley School of Business research paper, 18-87. https://doi.org/10.2139/ssrn.3269541

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

Pennycook, G., Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Peters, U. (2020). What is the function of confirmation bias? Erkenntnis, 87(3), 1351–1376. https://doi.org/10.1007/s10670-020-00252-1

Prakash, A., Verma, A., Sharma, Dr., Das, A. (2023). Disentangling the Effect of Confirmation Bias and Media Literacy on Social Media Users' Susceptibility to Fake News. Journal of Content, Community & Communication, 17(9), 16-30.

Quattrociocchi, W., Scala, A., &Sunstein, C. R. (2016). Echo Chambers on Facebook. Social Science Research Network. https://doi.org/10.2139/ssrn.2795110

R Core Team. (2022). R: A language and environment for statistical computing. Retrieved from https://www.r-project.org/. Accessed July 4, 2024.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in Ecology and Evolution, 11(9), 1141–1152. https://doi.org/10.1111/2041-210x.13434

AUTHOR, (2021). Retrieved from Link Accessed July 6 2024.

Sharot, T., & Garrett, N. (2016). Forming Beliefs: Why Valence matters. Trends in Cognitive Sciences, 20(1), 25–33. https://doi.org/10.1016/j.tics.2015.11.002

Simko, J., Hanakova, M., Racsko, P., Tomlein, M., Moro, R., &Bielikova, M. (2019). Fake news reading on social media: an eye-tracking study. In Proceedings of the 30th ACM conference on hypertext and social media (pp. 221-230). https://doi.org/10.1145/3342220.3343642

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

Sülflow, M., Schäfer, S., & Winter, S. (2018). Selective attention in the news feed: An eye-tracking study on the perception and selection of political news posts on Facebook. New Media & Society, 21(1), 168–190. https://doi.org/10.1177/1461444818791520

Tappin, B. M., van der Leer, L., & McKay, R. T. (2017). The heart trumps the head: Desirability bias in political belief revision.*Journal of Experimental Psychology: General, 146*(8), 1143–1149. https://doi.org/10.1037/xge0000298

van der Meer, T. G. L. A., & Brosius, A. (2024). Credibility and shareworthiness of negative news. Journalism, 25(1), 61-80. https://doi.org/10.1177/14648849221110283

Waddell, T. F. (2017). What does the crowd think? How online comments and popularity metrics affect news credibility and issue importance. New Media & Society, 20(8), 3068–3083. https://doi.org/10.1177/1461444817742905

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

Wickens, C. D., Hollands, J. G., Banbury, S., &Parasuraman, R. (2015). Engineering Psychology and human performance. In Psychology Press eBooks. https://doi.org/10.4324/9781315665177

Winter, S., Brückner, C., &Krämer, N. C. (2015). They came, they liked, they commented: Social influence on Facebook news channels. Cyberpsychology, Behavior, and Social Networking, 18(8), 431–436. https://doi.org/10.1089/cyber.2015.0005