Studia Psychologica, Vol. 67, No. 1, 2025, 54-68
https://doi.org/10.31577/sp.2025.01.911

The Higher You Think of Yourself, the Harder You Fall: Overconfidence as a Distinct, Mutable Predictor of Fake News Vulnerability

Eugen-Călin Secară, Nicolae-Adrian Opre
Department of Psychology, Faculty of Psychology and Education Sciences, Babeș-Bolyai University, Romania

This study investigated whether memory overconfidence is predictive of fake news vulnerability when controlling for previously established predictors and whether it can be experimentally reduced (*N* = 395). Participants completed measures of reflexive and reflective open-minded thinking, rated news articles collected from untrustworthy sources and completed general knowledge and working memory tasks. Confidence was assessed after each trial. The control group received feedback on the time spent on the tasks, while the experimental group was informed about the number of incorrect answers and their average confidence level. Afterwards, both groups completed the rest of the general knowledge and working memory tasks, alongside the confidence assessment and then rated other fake news articles. While neither reflexive nor reflective open-minded thinking significantly predicted fake news vulnerability, memory overconfidence did. Overconfidence correcting feedback reduced overconfidence but not fake news vulnerability. These findings indicate that memory overconfidence is a robust, mutable predictor of fake news vulnerability and highlight the need for more in-depth behavioral research.

Key words: misinformation, disinformation, working memory, general knowledge, reflexive open-minded thinking

Introduction

Fake news involves the manipulation of information, either through the creation of entirely false content or the distorted presentation of factual information (Aïmeur et al., 2023). In recent years, it has become a staple of modern crises (Hunt et al., 2020). In politics, fake news has been shown to create social divisions by promoting conspiracy theories and capitalizing on affectively charged topics (e.g., child trafficking by democrats in the 2016 Pizzagate conspiracy) to influence election results (Guess et al., 2018). Regarding the growing tensions between Russia and

Correspondence concerning this article should be addressed to Nicolae-Adrian Opre, Department of Psychology, Faculty of Psychology and Education Sciences, Babeș-Bolyai University, Mihail Kogălniceanu 1, Cluj-Napoca 400347, Romania. E-mail: Adrian.opre@ubbcluj.ro

Supplementary materials are available at https://journals.savba.sk/index.php/studiapsychologica/article/view/2227/883

Received October 30, 2023

Ukraine, fake news was employed to sow social unrest by portraying Ukraine's leadership as corrupt, Nazi sympathizers (EU vs. Disinfo, 2021). During the COVID-19 pandemic, health-related fake news has diminished trust in governments by downplaying the gravity of the situation (Lasco, 2020), endorsing conspiratorial reasons for the states' actions (Douglas, 2021), and endangering the population by promoting hoax prevention measures or treatments (Moreno-Castro et al., 2022). The later stages of the pandemic saw a new wave of fake news aimed at discrediting vaccines (Loomba et al., 2021). The issue of fake news encompasses three primary concerns: believing fake news, sharing fake news, and acting based on fake news. Among these, we identify belief in fake news as the foundational issue since it precipitates the others to act as a vulnerability factor. Hence, our discussion focuses on the tendency to consider fake news as credible (i.e., accurate and believable, Vu & Chen, 2023), which we refer to as "fake news vulnerability".

Considerable progress has been made in understanding the phenomenon (Levy & Ross, 2021; Pennycook & Rand, 2021), its predisposing factors (Bronstein et al., 2019; Pennycook & Rand, 2020) and their interaction with the digital environment (Moravec et al., 2018; Mosleh et al., 2021). Working within the dual process framework (Evans & Stanovich, 2013), Pennycook and Rand (2020) show that reflexive thinkers are more likely to invest unwarranted trust in political fake news than reflective thinkers. They characterize reflexive open-minded thinking as a tendency to uncritically accept a wide range of claims of dubious epistemic quality (i.e., generalized response bias) and take it to encompass overclaiming and bullshit receptivity. Few studies have explored how cognitive processes, like memory or higher-level failures in meta-cognition (e.g., overconfidence), contribute to fake news vulnerability (Rapp & Withall, 2024). The current study aims to investigate whether memory overconfidence is predictive of fake news vulnerability when controlling for previously established predictors and whether it can be experimentally reduced.

Overconfidence

Overconfidence, defined either as an overestimation of personal ability, performance, or chances of success (Moore & Healy, 2008) or as the investment of personal beliefs with excessive certainty (Soll & Klayman, 2004), has been considered the most ubiquitous and potentially calamitous problem related to decision-making (Plous, 1993). The differences between accuracy and confidence are explained by task skill rather than intelligence (Lichtenstein & Fischhoff, 1977), similar to the popular Dunning-Kruger effect (Kruger & Dunning, 1999), which predicts inflated self-evaluations at low levels of performance and diminished self-evaluations at high levels. Moreover, overconfidence in one's reasoning has been associated with belief in COVID-19 conspiracy theories (Vranic et al., 2022), and overconfidence in one's ability to detect fake news has been associated with more frequent visits to news sources of dubious quality (Lyons et al., 2021) and was unaffected by demographic variables (Dobbs et al., 2023). Hence, examining the role of overconfidence as a fake news vulnerability factor, in addition to reflective and reflexive open-minded thinking, can offer additional information about who is vulnerable to fake news. However, as overconfidence is considered a failure of metacognitive monitoring (Miller & Geraci, 2014), it is present in a large repertoire of assessments (e.g., driving ability, dating popularity). The following observations support the current choice of investigating the role of memory overconfidence in fake news vulnerability. First, false memories about fake news are common, especially when the fake news supports the person's beliefs (Murphy et al., 2019, Leon et al., 2023). Second, people display overconfidence in their false memories even after being warned that some of the stories were fabricated. Even after being explicitly told that the information they encountered was false, individuals who display low cognitive ability failed to adequately update their attitudes (De keersmaecker & Roets, 2017). Finally, two types of conflict can be detected when analyzing information: local conflicts and global conflicts. Local conflicts are detected when contradicting information is presented in the same item (e.g., same article, same website). Global conflicts are detected when the present information contradicts previously held knowledge. Memory overconfidence may interfere with both types of conflict detection as people are less likely to revisit previous information, either local or global, in which they are confident. Therefore, we hypothesized that overconfidence in verbal working memory (H1) and general knowledge (i.e., long-term memory, H2) are predictive of fake news vulnerability, controlling for other reflexive and reflective open-minded thinking. As Fleming and Lau (2014) distinguish between metacognitive sensitivity (the ability to distinguish between correct and incorrect responses, H1.1 and H2.1) and metacognitive bias (the overall level of confidence expressed in incorrect trials, H1.2 and H2.2), we tested the predictive power of both measures.

Reflective Open-minded Thinking

Reflectively open-minded thinking requires 1) the allocation of time and cognitive resources needed to detect conflicts (De Neys, 2014) and 2) the willingness to reflect on one's own biases (Baron, 1993). The Cognitive Reflection Test (CRT, Frederick, 2005), which presents logical problems where the intuitive response is incorrect and must be reconsidered to find the right answer, and the self-report scale developed by Stanovich and West (1997), termed Actively Open-minded Thinking (AOT) scale, are considered to be indicative of this dimension and have been negatively associated with trust in political fake news (Bronstein et al., 2019). In their investigation of health-related fake news, Scherer et al. (2021) found cognitive reflection to be a unique but inconsistent predictor of vulnerability as it was negatively associated with misinformation about cancer and vaccination, but not with misinformation about statins.

On the opposite side of the spectrum, reflexive open-minded thinkers uncritically endorse a wide spectrum of epistemically suspect beliefs, guided by gut feelings. Pennycook and Rand (2020) include bullshit receptivity and overclaiming as indicators of reflexive open-minded thinking, and show that both are positively associated with fake news vulnerability.

Bullshitting is defined by Frankfurt (2005) as an alternative epistemic stance besides telling the truth and lying, which is characterized by a lack of regard for truth (as opposed to the latter, which requires preoccupation with truth so that it can be manipulated or concealed). The aim of bullshitting is to impress rather than inform (Pennycook et al., 2015). As a measure of bullshit receptivity, Pennycook et al. (2015) asked participants to rate the profundity of randomly generated sentences, considering them as bullshit because the agent who produced them lacked a world-model, therefore was unpreoccupied with truth. The results indicated that those who considered the sentences profound professed an intuitive thinking style and were more likely to believe paranormal claims.

Overclaiming is a particular form of impression management (Paulhus, 1984) which implies exaggerating one's knowledge. Its assessment consists of rating familiarity with several things, some of which do not exist (Paulhus et al., 2003). The extent to which people report being familiar with the invented items is indicative of overclaiming. While some studies have operationalized overconfidence through overclaiming (Anderson et al., 2012), Bensch et al. (2019) show that the two constructs do not share a nomological network and that they add distinct variance over that of personality measures.

Reducing Fake News Vulnerability

The inattention-based account of misinformation sharing (Pennycook et al., 2021b) stipulates that people distribute fake news on social media because their behavior is controlled by alternative reinforcers (e.g., social validation) instead of caring about the accuracy of what they share. Presenting a nudge towards accuracy, such as asking people to rate the accuracy of a headline, has been shown to reduce fake news sharing in lab experiments and on Twitter (Pennycook et al., 2021b). Providing overconfidence-correcting feedback (e.g., informing people of their overconfidence scores) can be considered a type of accuracy nudge and has been proposed as a potential intervention to reduce trust in fake news (Lyons et al., 2021; Mirhoseini et al., 2023). As reductions in memory overconfidence could increase local and global conflict detection, we hypothesized that receiving overconfidence-correcting feedback will diminish fake news vulnerability (H3.1), overconfidence acting as a change mechanism (H3.2).

Method

The preregistration for the study can be accessed at https://osf.io/x2nzy. Open Science Framework data, materials and code can be accessed at: https://osf.io/n8fyp/.

Participants

An a priori power analysis (G*Power 3.1, Faul et al., 2007) suggested that a minimum of 395 participants would be required to detect a small effect size (f^2 = .02) for the increase in R^2 stipulated by H1 and H2, and a medium effect size (f^2 = .25) for H3.1 and H3.2 (α = ·05 and power = ·80). A total of 740 participants were recruited from the Babeș-Bolyai student population, who were offered extra course credits, and from the general Romanian population, using general social media posts and targeted posts in various social media groups. The study was completed by 395 participants (335 female, 56 male, 4 indicating other gender identities, M_{age} = 20.69, SD_{age} = 4.29), and according to the preregistered plan, only completers were included in the analyses. Of the 354 incomplete accounts, 196 (55.37%) dropped out after answering the demographic questions and 119 (33.62%) dropped out before completing the first post-test task. All participants who completed the study were offered a chance to win one of the ten vouchers (worth 20 Euros each). The inclusion criterion was being at least 18 years old. Recruitment efforts included weekly posting and reposting for the duration of the recruitment period. Participants who opted for the vouchers provided their name and their email address and were informed that multiple entries would result in exclusion from the lottery.

In terms of educational level, 1% of the sample reported having completed middle school, 78.5% had completed high school, 15.7% had obtained a bachelor’s degree and 4.8% had obtained a master’s degree.

Measures

Fake news vulnerability was assessed using twelve health-related news articles, eight collected from untrustworthy news outlets (Berezow, 2017) and four randomly generated. Previous research on these articles illustrated the lack of differences between the two categories of items in terms of scores, factorial structure, and association with relevant variables, suggesting that they are representative of health-related fake news (Secară, 2018, 2019). Participants were asked to rate the trustworthiness of each item on a 5-point scale ranging from 1 – "Very low level of trust" to 5 – "Very high level of trust" and were presented with the following statement: "A trustworthy article is defined as containing accurate and honest information that you consider you can rely on". One set of six items was presented at pre-test (T0) and another at post-test (T1), with the order of items presented at each time point being randomized. The sets were selected based on data from previous studies. We selected articles that received similar scores and displayed good internal consistency in groups of 6 (Cronbach's α = .86 and .80, see Secară, 2019). Acceptable internal consistency was observed for the data in this study (Cronbach's α = .76 and .74). A fake news vulnerability score at each moment was computed by averaging the ratings of the articles. Six health-related articles collected from trustworthy outlets were included, three at each time point, to mask the aim of the study and avoid artificial skepticism. Each article featured a title (e.g., How to cure tuberculosis naturally with vitamin C) and consisted of four paragraphs (about 400 words total): an introduction defining the key terms (e.g., tuberculosis and its treatment), a paragraph on the underlying theory (e.g., references to books and studies, Vitamin C is the fuel for the body's own immune system), another detailing the specifics of the intervention (e.g., how much and how often), and a conclusion.

Pseudo-profound bullshit receptivity was measured using 10 items from the Bullshit Receptivity Scale (BRS, Pennycook et al., 2015). Participants were presented with 10 profound sounding randomly generated sentences (e.g., "Wholeness quiets infinite phenomena") and asked to rate their perceived profundity on a 4-point scale ranging from "not at all profound" to "very profound". Pseudo-profound bullshit receptivity was computed by averaging the profoundness scores accorded to the items of the BRS. Internal consistency was acceptable (Cronbach's α = .75).

Overclaiming was measured using the "historical names and events" and the "topics in physical sciences" subscales of the Over-Claiming Questionnaire (OCQ, Paulhus et al., 2003). Fifteen items from each category were presented (e.g., "centripetal force"), three of them being fictional (e.g., "ultra-lipid"). Participants were asked to rate their familiarity with the items on a scale ranging from 1 – "slightly familiar" to 6 – "very familiar" with the option 0 – "never heard of it". An overclaiming bias score was computed by summing the familiarity scores of the fabricated items. Internal consistency was good (Cronbach's α = .86).

Analytic thinking was measured using the Cognitive Reflection Test (CRT, Toplak, West, & Stanovich, 2014). Participants were presented with seven logical world problems that cue an incorrect intuitive response which needs to be examined and disregarded in order to reach the correct answer. The total CRT score was obtained by counting the number of correctly solved problems.

Actively open-minded thinking was measured using the 17-item version of the AOT scale developed by Svedholm-Häkkinen & Lindeman (2017). The scale assesses the disposition to think reflectively by asking participants to rate statements such as "People should always take into consideration evidence that goes against their beliefs" or "I consider myself broad-minded and tolerant of other people's lifestyles" on a scale from ranging from 1 – "strongly disagree" to 6 – "strongly agree". A general actively open-minded thinking score was obtained by summing the scores of the items. Internal consistency was acceptable (Cronbach's $\alpha = .74$).

Overconfidence related to verbal working memory was assessed using the operation span task (Unsworth et al., 2005). Each trial included a math equation which had to be solved as quickly and accurately as possible, followed by a word which needed to be remembered. The task includes three types of practice trials: equations only, words only and mixed. After three to seven experimental trials, participants were asked to recall the words in the correct order and assess how confident they are that their response is correct (on a scale from 0 to 100).

Overconfidence related to general information was assessed using the information subscale of the Multidimensional Aptitude Battery-II (MAB-II, Jackson, 1998; Iliescu, Glință, & Ispas, 2009). Participants were presented with 26 multiple choice questions related to general knowledge. After each question, they were asked to evaluate how confident they are that their response is correct (on a scale from 0 to 10).

Two metacognitive sensitivity scores were computed for both the verbal working memory and the general knowledge tasks. The first one was computed by subtracting the sum of the confidence ratings of the correct answers from sum of the confidence ratings of the incorrect answers in the first half of each test (pretest, T0) and the second, using the same formula, from the second half of each test (post-test, T1). Two metacognitive bias scores were computed for both the verbal working memory and the general knowledge tasks. The first was computed by averaging confidence ratings across the first half of the trials (pretest, T0) and the second by averaging confidence ratings across the second half of the trials (post-test, T1).

Procedure

Participants entered the study via a link to the Gorilla online experimental platform (Anwyl-Irvine et al., 2020). There they were presented with information about the study and had to give their consent before proceeding. As revealing the aim of the study might alter their responses (e.g., increase skepticism towards the news articles), they were told that the main aim of the study was to analyze how individual differences influence the response to certain types of feedback in memory tasks. After completing demographic information, participants completed the CRT, OCQ, BRS and AOT. They then rated half of the news items and completed half of the general knowledge and working memory tasks. After each trial, confidence was assessed using a slider. At this point, the platform randomly assigned participants to either the control or experimental group. Both groups received feedback, the control group on the time taken to complete the two tasks and the intervention group on the number of incorrect answers for each task and their average confidence in the incorrect answers. To ensure that the information was retained, participants had to type in the numbers presented in order to continue. They then completed the remainder of the general knowledge and working memory tasks, each trial followed by the confidence assessment, and then rated the remaining news articles. Once all the data had been collected, participants received an email informing them of the fake articles and their average trust scores for the reliable and unreliable articles.

Data Analysis

Hierarchical regressions were used to evaluate the predictive power of overconfidence, the criterion being fake news vulnerability at pre-test (T0). The base model included pseudo-profound bullshit receptivity, overclaiming, actively open-minded thinking, and analytical thinking as predictors. In the second model, the hypothesis specific predictors (e.g., sensitivity or bias for working memory or general knowledge overconfidence at T0) were added. Each variable included as a predictor was mean-centered to avoid multicollinearity.

A one-way ANCOVA was used to test H3.1, having group as a between-subjects factor, trust in health-related fake news at T0 as a covariate, and trust in health-related fake news at T1 as the outcome.

All data were analyzed using IBM SPSS 25.

Results

The exploratory analysis of the correlations between health-related fake news vulnerability and the previously established predictors of fake news vulnerability found no significant associations when controlling for multiple comparisons (Table 1).

As a first step in assessing the added variance explained by overconfidence, we analyzed the predictive power of reflective and reflexive open-minded thinking (Model 0, see Table 3). The overall model was not statistically significant [$F(4, 390) = 1.68$, $R^2 = .02$, $p = .153$]. None of the variables predicted trust in health-related fake news.

Afterwards, we added the T0 overconfidence variables in separate models. Each model controlled for the variables in Model 0 and was independent of all other models. The metacognitive overconfidence bias measures were predictive of fake news vulnerability in the case of verbal working memory and general knowledge (H1.2: $B = .03$, $SE = .01$, $p = .008$ and respectively H2.2: $B = .29$, $SE = .10$, $p = .004$), while metacognitive sensitivity measures were not (H1.1, H2.1, see Table 4).

To analyze changes in fake news vulnerability between the experimental and control groups we performed a one-way ANCOVA, having group as an independent variable, fake news vulnerability at T1 as the dependent variable, and fake news vulnerability at T0 as

Table 1 *Descriptive statistics and zero-order correlations for reflexive and reflective open-minded thinking, fake news vulnerability, and overconfidence at pretest*

	M	*SD*	1	2	3	4	5	6	7	8	9
1. CRT	3.30	2.44	—								
2. OCQ	5.98	6.02	-.02	—							
3. BSR	2.40	0.56	.03	.14	—						
4. AOT	72.78	9.26	-.01	-.07	.14	—					
5. MSM1	-74.37	22.89	.05	-.04	.05	-.05	—				
6. MSO1	-345.04	336.09	-.04	.02	.06	.00	.09	—			
7. MBM1	4.45	2.34	.06	.07	.09	-.11	.20	.03	—		
8. MBO1	58.07	25.22	.02	.14	.15	-.09	-.03	.15	.25	—	
9. FN1	2.54	0.79	.05	.03	.11	-.07	.09	.06	.16	.13	—

Note. CRT = Cognitive Reflection Test, OCQ = Overclaiming Questionnaire, BSR = Bullshit Receptivity, AOT = Actively Open-minded Thinking, MSM1 = Metacognitive Sensitivity General Knowledge at pretest, MSO1 = Metacognitive Sensitivity Working Memory at pretest, MBM1 = Metacognitive Bias General Knowledge at pretest, MBO1 = Metacognitive Bias Working Memory at pretest, FN1 = Fake News vulnerability at pretest

Table 2 *Means and standard deviations used for group comparisons*

	Group					
	Experimental *N* = 188		Control *N* = 207			
	M	*SD*	*M*	*SD*	*t*	*p*
FN1	2.49	0.78	2.59	0.81	1.27	0.204
MSM1	-75.43	22.52	-73.41	23.24	0.88	0.382
MSO1	-366.53	331.94	-325.72	339.41	1.20	0.230
MBM1	4.69	2.45	4.24	2.21	-1.94	0.053
MBO1	58.24	26.53	57.91	24.03	-0.13	0.898
FN2	2.40	0.74	2.50	0.77	1.36	0.175
MSM2	-41.36	32.62	-42.02	31.75	-0.20	0.838
MSO2	-648.23	256.03	-577.84	329.32	2.38	0.018
MBM2	5.63	2.19	5.68	2.08	0.24	0.811
MBO2	39.80	32.31	47.11	32.14	2.25	0.025

Note. FN1 = Fake News vulnerability at pretest. MSM1 = Metacognitive Sensitivity General Knowledge at pretest, MSO1 = Metacognitive Sensitivity Working Memory at pretest, MBM1 = Metacognitive Bias General Knowledge at pretest, MBO1 = Metacognitive Bias Working Memory at pretest, FN2 = Fake News vulnerability at post-test. MSM2 = Metacognitive Sensitivity General Knowledge at post-test, MSO2 = Metacognitive Sensitivity Working Memory at post-test, MBM2 = Metacognitive Bias General Knowledge at post-test, MBO2 = Metacognitive Bias Working Memory at post-test.

Table 3 *Model 0: Multiple regression analysis for variables predicting health-related fake news vulnerability*

Variable	*B*	SE_B	*β*	*p*
Constant	15.26	.24		< .001
Cognitive reflection	.10	.10	.05	.322
Overclaiming	.01	.04	.01	.841
Bullshit receptivity	.08	.04	.09	.069
Actively open-minded thinking	-.03	.03	-.06	.252

Table 4 *Predictive power of overconfidence in working memory and general knowledge on fake news vulnerability at T0*

	Verbal Working Memory				General Knowledge			
Model specific overconfidence variable	Metacognitive sensitivity Model 1.1		Metacognitive bias Model 1.2		Metacognitive sensitivity Model 2.1		Metacognitive bias Model 2.2	
	B	*β*	*B*	*β*	*B*	*β*	*B*	*β*
Constant	15.26**		15.26**		15.26**		15.26**	
Cognitive reflection	.10	.05	.09	.05	.09	.05	.08	.04
Overclaiming	.01	.01	-.01	-.01	.01	.02	.00	.00
Bullshit receptivity	.07	.09	.06	.08	.08	.09	.07	.08
Actively open-minded thinking	-.03	-.06	-.03	-.05	-.03	-.05	-.02	-.05
Model specific overconfidence variable (T0)	.00	.06	.04**	.14	.02	.08	.05**	.15
R^2	.02		.04		.02		.04	
F	1.51		2.79**		1.90		3.06**	
ΔR^2	.003		.018		.007		.021	
ΔF	1.28		7.10**		2.75		8.42**	

Note. Each model tests the associated hypothesis. * $p < .05$, ** $p < .01$

a covariate. The results showed that the model was not statistically significant [$F(2, 392) = 1.84$, $p = .161$, *partial* $\eta^2 = .01$], indicating that the feedback received by the two groups did not affect their assessment of the fake news articles or that the study lacked the statistical power to detect the effect of the feedback (H3.1). Regardless, the preregistered pairwise comparisons and the mediation analysis were not applicable (H3.2).

Exploratory pairwise comparisons (Table 2) indicate that the only significant between-groups differences can be seen at post-test in working memory sensitivity and bias ($t = 2.38$, $p = .018$ and $t = 2.25$, $p = .025$).

Discussion

The current study aimed to investigate the role of memory overconfidence in predicting health-related fake news vulnerability and how trust in fake news is influenced after being presented with overconfidence-correcting feedback. Measures of reflexive and reflective open-minded thinking, which have been established as predictors of political fake news (Pennycook & Rand, 2020), were included in the analysis to test whether they were also predictive of health-related fake news and whether memory overconfidence explained unique variance beyond that accounted for by these variables.

None of the reflective and reflexive open-minded thinking constructs reached statistical significance as predictors of health-related fake news vulnerability (Table 3). Given that the randomly generated pseudo-profound items were constructed starting from a database of tweets by Deepak Chopra (Pennycook et al., 2015), a proponent of holistic, alternative medicine, they share a conceptual basis with the fake news articles selected for this study and use similar language. The absence of a relationship is therefore surprising. The results showing no relationship between cognitive reflection and fake news vulnerability were also surprising given the findings of Scherer et al. (2020) and Pennycook and Rand (2020). However, Mustață et al. (2023) found a similar absence of effect in their investigation of security and defense fake news vulnerability in Central and Eastern Europe. Similar to studies of Western populations, they found an association between fake news vulnerability and actively open-minded thinking, which was not found in the current study.

It is important to note that our sample, which is predominantly young, female, and made up of undergraduates, is not nationally representative. According to a UK study by King and Greene (2024), being female increases vulnerability to health-related fake news, while education does not predict trust in fake news. Similarly, Arin et al. (2023) found that female participants in the UK were more vulnerable to political fake news, in contrast to their German counterparts, where education was predictive but gender was not. The demographic of well-educated, young, female university students could explain some of the differences observed in our study. These findings highlight the need for further cross-cultural studies to explore other factors that might influence these outcomes.

Both measures of metacognitive bias were predictive of fake news vulnerability (*B* = .04, *SE* = .02, *p* = .008 for verbal working memory and *B* = .05, *SE* = .02, *p* = .004 for general knowledge). While the effect size detected was small, our preregistered analysis accounted for this possibility. However, neither metacognitive sensitivity regarding verbal working memory nor that regarding general knowledge proved predictive of fake news vulnerability. Performance scores or summed confidence scores for correct and incorrect answers were not predictive of fake news vulnerability (see online Supplementary materials). The observed predictive power of only the metacognitive bias measures suggests that the pivotal factor is not merely the gap between an individual's confidence and accuracy. Instead, it is a more general tendency to exhibit overconfidence, irrespective of the actual correctness of the responses. Essentially, those who don't acknowledge their errors are more susceptible to fake news, regardless of the confidence they place in their correct answers. This finding aligns with previous literature indicating that overconfidence in one's reasoning and abilities is predictive of susceptibility to misinformation (Lyons et al., 2021; Vranic et al., 2022).

While we can confidently state that overconfidence in memory predicts health-related fake news vulnerability, the specific mechanisms underlying this relationship warrant further examination. We started from the hypotheses that overconfidence in working memory will reduce responses typically seen when encountering local (i.e., intra-text) conflict, e.g., returning to previous paragraphs, and that overconfidence in general knowledge will produce fewer responses typical of detecting global conflict (i.e., between information presented in the text and previously held knowledge), e.g., verifying aspects that are uncertain. Although the current research design cannot provide the specific evidence needed to validate these hypotheses, the findings are promising. They highlight the need for further research into the relationship between memory overconfidence and fake news vulnerability. One strategy would be to employ eye-tracking software to observe whether participants exhibiting working memory overconfidence will display fewer back-and-forth movements, suggesting a lack of local conflict detection. Another strategy would be to provide a search function while reading the articles and inform participants that they can search for information about which they are unsure. In such a design, we would expect that participants with increased general knowledge overconfidence would be less likely to use the search function. These designs can address another limitation of the current study, namely the articles used to measure fake news vulnerability are complex text and we relied on general, overarching assessments, without knowing how participants related to the different paragraphs. Analyzing the contextual factors that trigger (dis)trust (e.g., skipping to the end of the article) could further advance our understanding of fake news and how to combat its influence.

The experimental part of the current research aimed to determine whether feedback could reduce overconfidence and fake news vulnerability. After the initial tasks, the control group received feedback on the time they took to complete the working memory and general knowledge tasks. The experimental group was informed of their correct and incorrect answers and the summed confidence ratings of each. Regardless of the type of feedback received, participants' belief in fake news remained unchanged. Significant between-groups differences were observed on both measures of working memory overconfidence, with the experimental group showing less metacognitive bias and more sensitivity. This suggests that the experimental manipulation was effective in reducing working memory overconfidence.

Our intervention was similar to the one proposed by Lyons et al. (2021). However, we diverged in a key area: our feedback targeted memory overconfidence rather than overconfidence in fake news detection. Based on their model, one might predict a context-specific decrease in susceptibility to fake news; that is, when participants are made aware of the potential presence of fake articles, they're likely to approach subsequent articles with increased skepticism. However, our data do not confirm a direct causal relationship between memory overconfidence and fake news vulnerability. It is possible that the feedback given was not strong enough to transfer the reduction in overconfidence from memory tasks to the context of fake news articles. The previously suggested eye-tracking approach could provide molecular evidence regarding the extent of feedback transfer between tasks. This would be particularly informative if paired with different types of feedback, including feedback directly related to the news articles.

Despite evidence from previous research suggesting that accuracy nudges can reduce misinformation discernment and sharing (Pennycook et al., 2021b; Mirhoseini et al., 2023), our study did not replicate this effect. The inattention-based account of misinformation sharing suggests that different aspects of the hypercomplex social media environment control sharing behavior, irrespective of how accurately the news is perceived. Given our experimental setup, we anticipated minimal interference from such factors. Future studies should explore whether feedback that corrects overconfidence can serve as an effective accuracy nudge in social media environments.

According to the motivated system 2 reasoning (Kahan, 2016) approach, the reason people avoid revisiting certain information may not be due to a failure in detecting conflicts, but rather a reluctance to engage further with that specific information. This may account for the changes in overconfidence, but not in fake news vulnerability, as observed in the current study. Including measures of belief in complementary and alternative medicine could provide preliminary evidence. A comprehensive approach would also analyze physiological responses to various text segments (e.g., using the previously discussed methods), complemented by qualitative research on participants' existing beliefs about the content. Together, these methods would provide nuanced evidence for the debate between the classical reasoning model and motivated system 2 reasoning.

Our results should be interpreted with caution because of certain limitations. First, all study materials were presented in Romanian. As a result, cultural cognition, viewed as a form of motivated system 2 reasoning (Kahan, 2016; Mustață et al., 2023) and particular to post-communist countries, may have influenced our outcomes. However, studies from Ukraine, another post-communist country, have shown alignment with the existing literature (Erlich et al., 2022). This suggests that such cultural specificities may not have significantly influenced our findings. Nevertheless, the use of a convenience sample requires added caution, as it restricts the extrapolation of our findings to the broader Romanian population. Given the novelty of the aspects investigated and the experimental nature of the design, we consider the results to be relevant.

The current study found that memory overconfidence is a robust, experimentally mutable (i.e., changeable) predictor of fake news vulnerability when assessed alongside established predictors, opening new fake news research directions. Causal relationships and underlying mechanisms need to be further explored in different contexts.

Authors' ORCID

Eugen-Călin Secară
https://orcid.org/0000-0002-7040-5340
Nicolae-Adrian Opre
https://orcid.org/0000-0002-4041-0659

References

Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining*, *13*(1), 30. https://doi.org/10.1007/s13278-023-01028-5

Arin, K. P., Mazrekaj, D., & Thum, M. (2023). Ability of detecting and willingness to share fake news. *Scientific Reports*, *13*(1), 7298. https://doi.org/10.1038/s41598-023-34402-6

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of Personality and Social Psychology*, *103*(4), 718–35. https://doi.org/10.1037/a0029395

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Baron, J. (1993). Heuristics and biases in equity judgments: A utilitarian approach. In B. A. Mellers & J. Baron (Eds.), *Psychological perspectives on justice: Theory and applications* (pp· 109–137)· Cambridge University Press· https://doi.org/10.1017/CBO9780511552069.007

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2019). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment*, *26*(3), 351–363. https://doi.org/10.1177/1073191117700268

Berezow, A. (2017, October 17). *Infographic: The best and worst science news sites*. https://www.acsh.org/news/2017/03/05/infographic-best-and-worst-science-news-sites-10948

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/10.1037/h0101832

De keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, *65*, 107–110. https://doi.org/10.1016/j.intell.2017.10.005

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*(2), 169–187. https://doi.org/10.1080/13546783.2013.854725

Dobbs, M., DeGutis, J., Morales, J., Joseph, K., & Swire-Thompson, B. (2023). Democrats are better than Republicans at discerning true and false news but do not have better metacognitive awareness. *Communications Psychology*, *1*(1), 46. https://doi.org/10.1038/s44271-023-00040-x

Douglas, K. M. (2021). COVID-19 conspiracy theories. *Group Processes & Intergroup Relations*, *24*(2), 270–275. https://doi.org/10.1177/136843022098206

Erlich, A., Garner, C., Pennycook, G., & Rand, D. G. (2021). Does analytic thinking insulate against pro-Kremlin disinformation? Evidence from Ukraine. *Political Psychology*. https://doi.org/10.1111/pops.12819

EU Vs. Disinfo. (2021). Disinformation build up: ProKremlin media reinvigorate their focus on Ukraine. *EU Vs. Disinfo Disinformation Review, 238*. https://euvsdisinfo.eu/disinformation-build-up-pro-kremlin-media-reinvigorate-their-focus-on-ukraine/

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/174569161246068

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. https://doi.org/10.3758/BF03193146

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443. https://doi.org/10.3389/fnhum.2014.00443

Frankfurt, H. G. (2005). On bullshit. In *On Bullshit*. Princeton University Press. https://doi.org/10.2307/j.ctt7t4wr

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

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. *European Research Council*, *9*(3), 4.

Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin*, *31*(9), 1250–1263. https://doi.org/10.1177/0146167205274694

Hunt, K., Agarwal, P., Al Aziz, R., & Zhuang, J. (2020). Fighting fake news during disasters. *OR/MS Today*, *47*(1), 34–39. https://doi.org/10.1287/orms.2020.01.06

Iliescu, D., Glință, F., & Ispas, D. (2009). Cultural adaptation of MAB-II (Multidimensional Aptitude Battery) in Romania. *Psihologia Resurselor Umane*, *7*(1), 75–87. https://doi.org/10.24837/pru.v7i1.403

Jackson, D. N. (1984). *Multidimensional Aptitude Battery – Manual.* London: Research Psychologists Press.

Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. A. Scott & S. M. Kossly (Eds.*), Emerging trends in the social and behavioral sciences* (pp. 1–16). Wiley. https://doi.org/10.1002/9781118900772.etrds0417

King, N., & Greene, C. M. (2024). Susceptibility to cancer misinformation: Predictors of false belief and false memory formation. *Applied Cognitive Psychology*, *38*(1), e4184. https://doi.org/10.1002/acp.4184

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121. https://doi.org/10.1037/0022-3514.77.6.1121

Lasco, G. (2020). Medical populism and the COVID-19 pandemic. *Global Public Health*, *15*(10), 1417–1429. https://doi.org/10.1080/17441692.2020.1807581

Leon, C. S., Bonilla, M., Brusco, L. I., Forcato, C., & Benítez, F. U. (2023). Fake news and false memory formation in the psychology debate. *IBRO Neuroscience Reports*, *15*, 24–30. https://doi.org/10.1016/j.ibneur.2023.06.002

Levy, N., & Ross, R. M. (2021). The cognitive science of fake news. In *The Routledge handbook of political epistemology* (pp. 181–191). Routledge. https://doi.org/10.4324/9780429326769

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0

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*(3), 337–348. https://doi.org/10.1038/s41562-021-01056-1

Lyons, B. A., Montgomery, J. M., Guess, A. M., Nyhan, B., & Reifler, J. (2021). Overconfidence in news judgments is associated with false news susceptibility. *Proceedings of the National Academy of Sciences*, *118*(23), e2019527118. https://doi.org/10.1073/pnas.2019527118

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition*, *29*, 131–140. https://doi.org/10.1016/j.concog.2014.08.008

Mirhoseini, M., Early, S., El Shamy, N., & Hassanein, K. (2023). Actively open-minded thinking is key to combating fake news: A multimethod study. *Information & Management*, *60*(3), 103761. https://doi.org/10.1016/j.im.2023.103761

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Moreno-Castro, C., Vengut-Climent, E., Cano-Orón, L., & Mendoza-Poudereux, I. (2022). Exploratory study of the hoaxes spread via WhatsApp in Spain to prevent and/or cure COVID-19. *Gaceta sanitaria*, *35*, 534–540. https://doi.org/10.1016/j.gaceta.2020.07.008

Murphy, G., Loftus, E. F., Grady, R. H., Levine, L. J., & Greene, C. M. (2019). False memories for fake news during Ireland's abortion referendum. *Psychological Science*, *30*(10), 1449–1459. https://doi.org/10.1177/0956797620923299

Mustață, M. A., Răpan, I., Dumitrescu, L., Dobreva, H., Dimov, P., Andrzej, L., ... & Buță, C. (2023). Assessing the truthfulness of security and defence news in Central and Eastern Europe: The role of cognitive style and the promise of epistemic sophistication. *Applied Cognitive Psychology*, *37*(6), 1384–1396*.* https://doi.org/10.1002/acp.4130

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021a). A practical guide to doing behavioural research on fake news and misinformation. *Collabra, 7*, 25293. https://doi.org/10.1525/collabra.25293

Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2015). On the reception and detection of pseudo-profound bullshit. *Judgment and Decision Making*, *10*(6), 549–563. https://doi.org/10.1017/S1930297500006999

Pennycook, G.*,* Epstein, Z*.*, Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021b). Shifting attention to accuracy can reduce misinformation online. *Nature, 592*, 590–595. https://doi.org/10.1038/s41586-021-03344-2

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

Rapp, D. N., & Withall, M. M. (2024). Confidence as a metacognitive contributor to and consequence of misinformation experiences. *Current Opinion in Psychology, 55*, 101735. https://doi.org/10.1016/j.copsyc.2023.101735

Scherer, L. D., McPhetres, J., Pennycook, G., Kempe, A., Allen, L. A., Knoepke, C. E., ... & Matlock, D. D. (2021). Who is susceptible to online health misinformation? A test of four psychosocial hypotheses. *Health Psychology*, *40*(4), 274–284. https://doi.org/10.1037/hea0000978

Secară (2018). *Understanding and assessing bullshit vulnerability*. [Master Thesis, University of Vienna].

Secară (2019). *Măsurarea tulburărilor informaționale referitoare la domeniul sănătății. Delimitări conceptuale și mecanisme comune implicate în psihopatologie* (Measurement of health-related information disorders. Conceptual clarifications and common mechanisms involved in psychopathology). [Master Thesis, Babeș-Bolyai University].

Simons, D. J. (2013). Unskilled and optimistic: Overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review*, *20*(3), 601–607. https://doi.org/10.3758/s13423-013-0379-2

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357. https://doi.org/10.1037/0022-0663.89.2.342

Svedholm-Häkkinen, A. M., & Lindeman, M. (2018). Actively open-minded thinking: Development of a shortened scale and disentangling attitudes towards knowledge and people. *Thinking & Reasoning*, *24*(1), 21–40. https://doi.org/10.1080/13546783.2017.1378723

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Vranic, A., Hromatko, I., & Tonković, M. (2022). "I did my own research": Overconfidence, (dis)trust in science, and endorsement of conspiracy theories. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.931865

Vu, H. T., & Chen, Y. (2023). What influences audience susceptibility to fake health news: An experimental study using a dual model of information processing in credibility assessment. *Health Communication, 39*(6), 1113–1126. https://doi.org/10.1080/10410236.2023.2206177