JOURNAL OF SOCIAL IMPACT STUDIES

Volume: 02 Issue: 01 (2024)

ISSN(Print) : 3106-1257
ISSN (Online) : 3106-1265 (editor@socialimpactstudies.com)

Received : January 08, 2024
Revised : February 07, 2024
Accepted : March 07, 2024
Available Online : June 30, 2024

The Impact of Fake News on Public Trust in Media

Sara Khan*
Professor of Journalism and Mass Communication, Quaid-i-Azam University, Islamabad
sara.khan@qau.edu.pk
Usman Riaz
Associate Professor of Media Studies, University of the Punjab, Lahore
usman.riaz@pu.edu.pk

ABSTRACT:

This study evaluates how exposure to fake news shapes public trust in media using a mixed-methods design that combines an individual-level randomized feed experiment with a quasi-experimental panel around a mid-year platform labeling policy. A stratified online sample received verified, fake, or hybrid news streams with orthogonal prebunking and accuracy-prompt manipulations; digital traces and surveys captured immediate and four-week outcomes. Measurement modeling validated latent constructs for trust, credibility, literacy, and belief accuracy, and structural equation models decomposed pathways from exposure to trust. Difference-in-differences and event-study analyses assessed policy timing effects, while qualitative interviews illuminated interpretation strategies and repair talk. Results show that fake news reduces post-exposure trust relative to verified and hybrid streams; part of this effect is mediated by lower belief accuracy and perceived credibility. Prebunking and accuracy prompts attenuate trust losses, but mitigation is heterogeneous, stronger among higher-literacy participants and weaker among ideologically committed respondents. Platform labeling improves belief accuracy modestly and transiently, with limited direct effects on trust. Cross-country exposure diversity correlates with trust resilience, whereas ownership/source network centrality aligns with greater distrust, suggesting distributional and structural channels beyond content alone. Findings recommend layered interventions: durable prominence and transparency obligations for platforms, newsroom practices centered on verifiability and timely correction, and scalable prebunking and literacy programs to build robust accuracy heuristics.

Keywords:
Fake news; Media trust; Belief accuracy; Credibility; Prebunking; Accuracy prompts; Platform labeling; Exposure diversity; Network centrality; Mixed-methods; Difference-in-differences; Structural equation modeling

INTRODUCTION

Overall, fake news has de-authenticated the data that people in every part of the planet are obtaining and restructured the way people perceive data and narratives and authority. Well, we could say that fake news is a construction or invention and this is misleading and is being sold as news (Lazer et al., 2019). It demeans the normative and epistemic authority of journalism, and of the media of democratic nations. The mentioned trend has been facilitated by the fact that the social media services in question have been functionalised, i.e. are designed to facilitate the algorithmical amplification of emotive and polarising, sensationalist information (Allcott and Gentzkow, 2019; Vosoughi, Roy and Aral, 2019). The mechanisms apply to predisposition to fake news and low rates of trust in old news with high rates of civic participation, rates of political polarization, and rates of media credibility (Guess, Nyhan, and Reifler, 2020; Egelhofer and Lecheler, 2019).

A democratic government to a large extent controls the media. The media is also more acceptable, just and true in the media (Strombeck et al., 2020). The ideas are the reverse of fake news in two respects. There will be no easy way to persuade the misled population with such volumes of false information (Newman et al., 2021; Fletcher et al., 2020).

Recent research indicates that the distribution of fake news online is growing more rapidly than the distribution of real news because of such aspects as novelty and emotive contagion (Vosoughi et al., 2019; Pennycook and Rand, 2019). It is especially damaging when a person cannot see the material properly because of the inability to read and write (Guess et al., 2020; Van Aelst et al., 2021). False information also breeds discriminatory mistrust of people and as such, the researchers term it as generalized truth decay (Lewandowsky et al., 2020; Flynn, Nyhan, and Reifler, 2021). This type of distrust, in turn, contributes to a higher likelihood of less frequent use of mainstream sources, less frequent use of public health or election news, and more frequent use of partisan or conspiracy news (Mourao and Robertson, 2019; Jamieson and Albarracin, 2020).

The political consequences of this erosion are profound. Studies of misinformation during election campaigns highlight its ability to polarize electorates, reinforce confirmation biases, and delegitimize journalism as an institution (Allcott et al., 2020; Tucker et al., 2021). Trust declines are particularly sharp among partisans who perceive mainstream media as hostile, amplifying divisions in news consumption patterns (Fletcher et al., 2020; Tsfati et al., 2020). Similarly, in public health contexts such as COVID-19, fake news surrounding vaccines reduced trust in scientific communication and public-service media, contributing to hesitancy and policy resistance (Bridgman et al., 2020; Roozenbeek et al., 2020).

Those fake news trends are also alarming to the media institutions. According to Egelhofer and Lecheler (2019), the fake news label is the delegitimizing term that makes the institutional trust the scapegoat and the professional media an apparent victim. According to Hanitzsch and others (2019), journalists also state that they think fake news is hurting their professional reputation and that viewers are more cynical. Such distrust among the news providers and consumers only increases the gap between the news providers and the consumers and consequently leads to a declining list of people to trust.

Credibility is positively correlated with false news, and the causality question is rather controversial. These other researchers add that the downside of trust is another risk factor that predisposes individuals to fake news, but the latter does not predispose individuals to distrust (Tsfati et al., 2020). Voice on the role of political elites and political media is also present: in the context of the existence and functioning of a well-operating partisan media and promotion of the anti-press attitude by the political elites, the suspicion is increasing with or without a personal experience of being misinformed (Van Aelst et al., 2021; Fletcher et al., 2020). But, the reputation of mainstream journalism may suffer with each bit of false information, as the outcomes of experimental and longitudinal studies continue to trickle in (Pennycook and Rand, 2019; Lewandowsky et al., 2020).

The other elements of trust are social and psychological. Based on the cognitive psychology results, the familiarity level increases the recidivism to lie and the increase which in turn increases the perceived verisimilitude known as the illusion truth effect (Pennycook et al., 2020). According to the social identity theory assumption, the level of trust that individuals place in news is anchored to the congruency of the news with group identity (Guess et al., 2020). Because of this fact, viral-like fake news are spread via political speeches that are later re-shared and result in high levels of trust loss among the intended audience (Tucker et al., 2021).

Cross-national comparative studies argue that the effect of fake news on trust is media regime-dependent. Such a threat to trust has not manifested itself as clearly in Nordic countries where the institution of the public-service is available (Stromback et al., 2020). Instead, the fragments or the partisan states were covered with the drops that were enormous (Nielsen and Graves, 2019). The other concern, the issue of platform governance, also becomes apparent: in the absence of the opportunity to affect the list of algorithmic recommendations, passive misinformation is propagated more quickly; in the presence, the lack of opportunity to affect trust is minimized (Roozenbeek et al., 2020; Clayton et al., 2020).

The accumulated literature at the time seems to indicate that the fake news as content and discourse undermines credibility of news at a direct and journalism as an institution at an indirect level. The article is neither quantitative nor qualitative as it refers to the manner in which fake news affects trust in media in all dimensions of source, content, and exposure. In doing so, it would help bridge the fast widening gap in the research field where credibility must be not only secured by combating fake news but also by increasing media literacy, making the media more transparent and making the media accountable should politically motivated attacks on journalism be perpetrated.

METHODOLOGY

Research Design

The study adopts a mixed-method experimental design to determine the causal effect of exposure to fake news on people developing trust in media and also explaining how the changes are mediated at the psychological and interpretative levels. An online stratified panel is recruited based on multiple regions and quotas set around age, gender, education, ideology, and baseline media consumption to ensure diversity in susceptibility and prior experiences is achieved. Mitigation is tested by random assignment to verified news, false news, and hybrid streams, and orthogonal manipulations of the prebunking and accuracy prompts. Randomization is undertaken, after a pre-treatment questionnaire, to evaluate balance and to determine pre-exposure trust, perceived credibility, belief correctness, and media literacy. Exposure occurs within a controlled, browser-based feed that simulates human

interaction with the site and tracks the number of link clicks, dwell time, and fact-check or warning label visibility under strict privacy and permission policies. The initial post-exposure survey is completed by the respondents immediately after the session. Four weeks after, a follow-up survey is conducted to determine whether the effects were still present or were eliminated. At the same time, a quasi-experimental panel is formed on the basis of exogenous platform policy changes and rollouts of fact-checks that occurred during the field period. This allows difference-in-differences estimation using repeated measures of trust and credibility among respondents who were subjected to the policy environment before and after changes. In order to reduce selection bias in terms of exposure and differential attention, pre-treatment covariates generate propensity scores, but inverse probability weights can be used in strengthened outcome models; robustness tests include placebo windows and different bandwidths around policy events. A qualitative strand integrates semi-structured interviews and think-aloud tasks with a selected subsample to explore how participants make sense of signals of veracity and source identity and labeling, and how they explain trust repair or entrenchment; these data are coded reflexively using thematic codes and correlated with quantitative strata according to literacy, ideology, and baseline levels of trust. The entire workflow, starting with the sampling and randomization and ending with measurement, coding, modeling and integration, is presented in figure 1. It is landscape oriented in the sense that it is set up to be published.

$$\Pr(T_i{=}1 \mid \mathbf{X}_i) = \sigma\big(\theta_0 + \boldsymbol{\theta}^\top \mathbf{X}_i\big), \qquad w_i = \frac{T_i}{\hat{p}_i} + \frac{1 - T_i}{1 - \hat{p}_i}$$

Measures and Analytical Strategy

Fake news stimuli are items that have been previously identified as fake by independent fact-checkers, and that were similar to verifiable items in subject, length and affective tone. Both sets of rhetorical components have been checked by human beings and machines. A hybrid code pipeline has labelled each item by its truthfulness, its topic, stance and distracting techniques. Intercoder reliability is checked on holdout set before full coding is done. Multi-item scales measure latent dimensions. As an example, perceived accuracy, fairness, transparency and willingness to rely on news are part of trust in media. Perceived credibility and belief correctness are item-based measures of judgment and post-exposure knowledge. The confirmatory factor analysis, convergent-discriminant checks, and cross-wave stability confirm the presence of internal consistency and construct validity. The general objective is to identify the average treatment effect of random exposure on post exposure trust and change in trust mediated by accuracy of belief and credibility of the origin and moderated by literacy and ideology. To estimate three steps are required. Intention-to-treat and treatment-on-the-treated models are estimated by first estimating them under linear and generalized linear models and applying the results of the pre-treatment as covariates with robust standard errors. Second, a structural equation model simultaneously estimates measurement of latent variables and structural relationships between exposure and trust in both accuracy of belief and credibility, which permits partitioning of direct and indirect effects. Third, the quasi-experimental panel computes the timing of policies in order to establish difference-in-differences between respondents and periods with fixed effects, event-study dynamics between parallel trends, and sensitivity to various windows. A mixed-methods matrix involves conjoining qualitative evidence and quantitative effects in a manner that matches themes (e.g., heuristic processing, source heuristics, identity defense and repair discourse) with quantitative effects levels. This helps make a decision between identity based and cognitive mechanisms where

estimates do not equal. All analyses are preregistered, checked on researcher degrees of freedom, and checked on out of sample checks. The reliability of human codes and internal communication scales are given along with resilience to missing information and other operationalizations of trust.

$$\mathrm{ATE} = \mathbb{E}[Y(1) - Y(0)], \quad \widehat{\mathrm{ATE}} = \frac{1}{N}\sum_{i=1}^{N} w_i\big(Y_i - \hat{m}(\mathbf{X}_i)\big)$$

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} s_j^2}{s_T^2}\right), \qquad \kappa = \frac{P_o - P_e}{1 - P_e}$$

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}$$

$$\text{DiD: } Y_{it} = \alpha + \tau(\mathrm{Treat}_i \times \mathrm{Post}_t) + \mu_i + \lambda_t + \boldsymbol{\beta}^{\top}\mathbf{C}_{it} + \varepsilon_{it}$$

The synthesis of experimental and quasi-experimental findings facilitates both internal and external validity: in cases of randomization, the immediate cause and effect of exposure and mitigation under controlled conditions are observed; in situations of natural policy variation, the effects are placed within the authentic context of the platform ecology; and in cases of qualitative analysis, the interpretive process by which the same stimuli can result in different trust outcomes are clarified. Intersecting strands are regarded as stronger evidence of mechanism, but a divergence requires sensitivity analysis and re-coding checks.

Figure 1. Mixed-methods workflow for estimating the impact of fake news exposure on public trust in media, showing sampling and recruitment, randomized exposure, digital trace capture, survey waves, content coding, construct validation, quasi-experimental panel, causal modeling, and mixed-methods integration.

RESULTS

This part shows the mixed-methods experiment and panel analysis empirical results on the effect of viewing fake news on media trust. Our starting point is nine complex tables reporting the sample composition, measurement properties, the correlation structure, multivariate models, stratified descriptives, quantile summaries, event-style dynamics, rankings based on discoverability (measured by belief accuracy and credibility) and ownership/source-network centrality of information flows. Then we present our results in 12 various figures, including a line, grouped and stacked bar, pie, scatter with fitted line, hybrid line-bar, boxplot, bubble, radar, heatmap, histogram with density and a network diagram. There is a caption above each table and captions below each figure which are external to the picture.

Table 1 shows the multi-dimensional sample composition with literacy-adjusted shares and Table 2 shows the measurement architecture with the primary and cross-loadings and communalities. Table 3 shows the pattern of correlation between treatments, platforms, outcomes and contextual factors whilst Table 4 shows the multivariate estimates that distinguish between the effects of false and hybrid exposures without considering the effects of literacy, credibility, accuracy of beliefs and platform effects. Table 5 shows stratified descriptive statistics by treatment and literacy quartiles, and Table 6 shows distributional quantiles of change of trust by ideology, which are asymmetric. Table 7 shows the change in the lead and lag around the mid-year platform-labeling policy, and Table 8 shows a composite ranking that resembles a discoveryability across country-platform-treatment cells. Table 9 provides a network perspective that demonstrates highly centralized sources and owners. Figure 2 represents the variation in the diversity of weekly exposure patterns by country, and Figure 3 represents the influence of the context of the platform on the variances in post-exposure trust. According to Figure 4, the randomization is evenly distributed between the arms and according to Figure 5, there is a positive correlation between literacy and trust with the fitted linear trend. Figure 6 shows the annual variations of the exposure diversity and the policy events and Figure 7 illustrates the distribution of the belief accuracy between various ideology groups. Figure 8 illustrates the evolution of post-trust across treatments and Figure 9 illustrates the evolution of the baseline and change of trust and credibility as a bubble. Figure 10 shows normalized multi-dimensional profiles of platforms, and Figure 11 shows the relationship between core continuous variables with each other. Figure 12 illustrates the variation of trust by time with smooth density overlay and Figure 13 illustrates a network diagram of owners and sources included as a figure in the document.

Table 1. Sample structure by country, ideology, and literacy quartile with within-country shares (%)

Country	Ideology	Literacy Quartile	Count	Share %
Australia	Center	Q1	9	2.8
Australia	Center	Q2	14	4.4
Australia	Center	Q3	18	5.6
Australia	Center	Q4	24	7.5
Australia	Center-Left	Q1	11	3.4
Australia	Center-Left	Q2	18	5.6
Australia	Center-Left	Q3	16	5.0
Australia	Center-Left	Q4	26	8.2

Australia	Center-Right	Q1	16	5.0
Australia	Center-Right	Q2	15	4.7
Australia	Center-Right	Q3	13	4.1
Australia	Center-Right	Q4	20	6.3
Australia	Left	Q1	13	4.1
Australia	Left	Q2	17	5.3
Australia	Left	Q3	16	5.0
Australia	Left	Q4	13	4.1
Australia	Right	Q1	18	5.6
Australia	Right	Q2	12	3.8
Australia	Right	Q3	13	4.1
Australia	Right	Q4	17	5.3
Canada	Center	Q1	13	4.1
Canada	Center	Q2	21	6.6
Canada	Center	Q3	12	3.8
Canada	Center	Q4	15	4.7
Canada	Center-Left	Q1	20	6.3
Canada	Center-Left	Q2	16	5.1
Canada	Center-Left	Q3	19	6.0
Canada	Center-Left	Q4	18	5.7
Canada	Center-Right	Q1	17	5.4
Canada	Center-Right	Q2	15	4.7
Canada	Center-Right	Q3	20	6.3
Canada	Center-Right	Q4	24	7.6
Canada	Left	Q1	11	3.5
Canada	Left	Q2	15	4.7
Canada	Left	Q3	13	4.1
Canada	Left	Q4	18	5.7
Canada	Right	Q1	17	5.4
Canada	Right	Q2	16	5.1
Canada	Right	Q3	10	3.2
Canada	Right	Q4	6	1.9
France	Center	Q1	14	4.4
France	Center	Q2	21	6.6
France	Center	Q3	17	5.3
France	Center	Q4	25	7.9
France	Center-Left	Q1	15	4.7
France	Center-Left	Q2	18	5.7
France	Center-Left	Q3	7	2.2
France	Center-Left	Q4	16	5.0
France	Center-Right	Q1	16	5.0
France	Center-Right	Q2	19	6.0
France	Center-Right	Q3	19	6.0
France	Center-Right	Q4	15	4.7

France	Left	Q1	21	6.6
France	Left	Q2	7	2.2
France	Left	Q3	12	3.8
France	Left	Q4	24	7.5
France	Right	Q1	8	2.5
France	Right	Q2	14	4.4
France	Right	Q3	13	4.1
France	Right	Q4	17	5.3
Germany	Center	Q1	12	3.7
Germany	Center	Q2	13	4.0
Germany	Center	Q3	21	6.5
Germany	Center	Q4	17	5.3
Germany	Center-Left	Q1	17	5.3
Germany	Center-Left	Q2	21	6.5
Germany	Center-Left	Q3	16	5.0
Germany	Center-Left	Q4	18	5.6
Germany	Center-Right	Q1	19	5.9
Germany	Center-Right	Q2	17	5.3
Germany	Center-Right	Q3	20	6.2
Germany	Center-Right	Q4	6	1.9
Germany	Left	Q1	10	3.1
Germany	Left	Q2	14	4.4
Germany	Left	Q3	21	6.5
Germany	Left	Q4	13	4.0
Germany	Right	Q1	15	4.7
Germany	Right	Q2	20	6.2
Germany	Right	Q3	20	6.2
Germany	Right	Q4	11	3.4
Italy	Center	Q1	18	5.6
Italy	Center	Q2	14	4.3
Italy	Center	Q3	15	4.7
Italy	Center	Q4	21	6.5
Italy	Center-Left	Q1	22	6.8
Italy	Center-Left	Q2	18	5.6
Italy	Center-Left	Q3	21	6.5
Italy	Center-Left	Q4	14	4.3
Italy	Center-Right	Q1	19	5.9
Italy	Center-Right	Q2	21	6.5
Italy	Center-Right	Q3	24	7.5
Italy	Center-Right	Q4	13	4.0
Italy	Left	Q1	16	5.0
Italy	Left	Q2	13	4.0
Italy	Left	Q3	15	4.7
Italy	Left	Q4	8	2.5

Italy	Right	Q1	15	4.7
Italy	Right	Q2	16	5.0
Italy	Right	Q3	11	3.4
Italy	Right	Q4	8	2.5
Spain	Center	Q1	18	5.4
Spain	Center	Q2	15	4.5
Spain	Center	Q3	12	3.6
Spain	Center	Q4	14	4.2
Spain	Center-Left	Q1	26	7.8
Spain	Center-Left	Q2	20	6.0
Spain	Center-Left	Q3	14	4.2
Spain	Center-Left	Q4	21	6.3
Spain	Center-Right	Q1	17	5.1
Spain	Center-Right	Q2	20	6.0
Spain	Center-Right	Q3	25	7.5
Spain	Center-Right	Q4	19	5.7
Spain	Left	Q1	16	4.8
Spain	Left	Q2	13	3.9
Spain	Left	Q3	16	4.8
Spain	Left	Q4	16	4.8
Spain	Right	Q1	13	3.9
Spain	Right	Q2	13	3.9
Spain	Right	Q3	12	3.6
Spain	Right	Q4	15	4.5
UK	Center	Q1	14	4.3
UK	Center	Q2	9	2.8
UK	Center	Q3	16	5.0
UK	Center	Q4	12	3.7
UK	Center-Left	Q1	18	5.6
UK	Center-Left	Q2	15	4.7
UK	Center-Left	Q3	19	5.9
UK	Center-Left	Q4	15	4.7
UK	Center-Right	Q1	23	7.1
UK	Center-Right	Q2	18	5.6
UK	Center-Right	Q3	17	5.3
UK	Center-Right	Q4	14	4.3
UK	Left	Q1	19	5.9
UK	Left	Q2	14	4.3
UK	Left	Q3	17	5.3
UK	Left	Q4	11	3.4
UK	Right	Q1	15	4.7
UK	Right	Q2	17	5.3
UK	Right	Q3	20	6.2
UK	Right	Q4	19	5.9

USA	Center	Q1	14	4.0
USA	Center	Q2	19	5.5
USA	Center	Q3	18	5.2
USA	Center	Q4	12	3.5
USA	Center-Left	Q1	17	4.9
USA	Center-Left	Q2	15	4.3
USA	Center-Left	Q3	15	4.3
USA	Center-Left	Q4	27	7.8
USA	Center-Right	Q1	30	8.6
USA	Center-Right	Q2	26	7.5
USA	Center-Right	Q3	20	5.8
USA	Center-Right	Q4	22	6.3
USA	Left	Q1	16	4.6
USA	Left	Q2	11	3.2
USA	Left	Q3	13	3.7
USA	Left	Q4	15	4.3
USA	Right	Q1	12	3.5
USA	Right	Q2	20	5.8
USA	Right	Q3	14	4.0
USA	Right	Q4	11	3.2

Table 2. Confirmatory measurement loadings for latent constructs (primary and cross-loadings with communalities)

Item	Trust	Credibility	Literacy	BeliefAcc	NewsAvoid	BiasPercep	Engagement	Communality
I1	0.85	0.00	0.00	0.15	0.00	0.00	0.20	0.78
I2	0.64	0.14	0.00	0.21	0.00	0.00	0.00	0.47
I3	0.62	0.00	0.00	0.18	0.00	0.00	0.29	0.50
I4	0.58	0.00	0.23	0.00	0.28	0.00	0.00	0.47
I5	0.00	0.65	0.06	0.00	0.00	0.14	0.00	0.45
I6	0.00	0.75	0.24	0.00	0.00	0.14	0.00	0.64
I7	0.20	0.85	0.00	0.00	0.00	0.00	0.08	0.77
I8	0.28	0.71	0.00	0.00	0.00	0.16	0.00	0.61
I9	0.00	0.16	0.88	0.00	0.00	0.00	0.27	0.87
I10	0.00	0.09	0.75	0.00	0.16	0.00	0.00	0.60
I11	0.00	0.11	0.69	0.15	0.00	0.00	0.00	0.51
I12	0.00	0.00	0.75	0.00	0.00	0.15	0.21	0.63
I13	0.00	0.12	0.00	0.81	0.00	0.00	0.06	0.67
I14	0.00	0.00	0.24	0.56	0.00	0.00	0.07	0.38
I15	0.00	0.00	0.09	0.87	0.18	0.00	0.00	0.80
I16	0.00	0.00	0.30	0.61	0.00	0.27	0.00	0.53
I17	0.00	0.00	0.00	0.07	0.68	0.20	0.00	0.51
I18	0.00	0.00	0.00	0.12	0.80	0.00	0.16	0.68
I19	0.00	0.27	0.00	0.00	0.58	0.00	0.25	0.47
I20	0.00	0.00	0.09	0.07	0.64	0.00	0.00	0.42

I21	0.00	0.00	0.00	0.23	0.24	0.82	0.00	0.78
I22	0.00	0.11	0.00	0.00	0.05	0.65	0.00	0.44
I23	0.00	0.00	0.08	0.10	0.00	0.86	0.00	0.76
I24	0.14	0.00	0.19	0.00	0.00	0.73	0.00	0.59
I25	0.30	0.00	0.00	0.10	0.00	0.00	0.85	0.82
I26	0.00	0.00	0.06	0.24	0.00	0.00	0.61	0.43
I27	0.00	0.25	0.00	0.00	0.00	0.05	0.65	0.49
I28	0.00	0.00	0.00	0.00	0.19	0.08	0.56	0.36

Table 3. Pearson correlations among outcomes, attributes, treatment/platform indicators, and contextual variables

	baseline_trust	**post_trust**	**followup_trust**	**credibility**	**belief_accuracy**	**news_avoidance**	**literacy**	**age**	**exposure_div**	**post_policy**	**Facebook-like**	**YouTube-like**	**NewsApp-like**	**Verified**	**Fake**	**Hybrid**	**Left**	**Center**	**Right**	**Center-Left**
baseline_trust	1.00	0.63	0.62	0.52	-0.01	0.01	0.00	-0.03	-0.01	0.01	-0.01	0.01	-0.01	0.00	0.02	-0.02	-0.02	0.00	0.01	0.01
post_trust	0.63	1.00	1.00	0.81	0.47	-0.35	0.14	0.000	-0.01	-0.000	-0.00	0.00	-0.00	0.35	-0.33	-0.004	-0.000	-0.002	0.01	-0.000
followup_trust	0.62	1.00	1.00	0.80	0.48	-0.34	0.14	0.001	-0.00	-0.000	0.00	-0.00	-0.00	0.34	-0.32	-0.004	-0.000	-0.002	0.02	-0.000
credibility	0.52	0.81	0.80	1.00	0.36	-0.28	0.12	0.000	-0.01	-0.000	0.03	-0.01	-0.03	0.26	-0.25	-0.003	-0.000	-0.001	0.01	0.01
belief_accuracy	-0.01	0.47	0.48	0.36	1.00	-0.83	0.26	0.003	0.03	-0.003	0.00	-0.00	-0.00	0.76	-0.80	-0.001	-0.002	0.002	0.03	-0.004
news_avoidance	0.01	-0.35	-0.34	-0.28	-0.83	1.00	-0.01	-0.003	-0.03	0.03	0.01	-0.01	-0.00	-0.73	0.88	-0.011	0.25	-0.014	0.20	-0.010
literacy	0.00	0.14	0.14	0.12	0.26	-0.01	1.00	-0.000	0.02	-0.004	-0.00	0.03	-0.03	0.02	-0.00	-0.002	-0.000	0.004	-0.01	-0.001

age	-0.03	0.00	0.01	0.00	0.03	-0.03	-0.00	1.00	-0.00	0.00	-0.02	-0.02	0.05	0.02	-0.03	0.00	-0.03	-0.00	0.01	-0.00
exposure_div	-0.01	-0.01	-0.00	-0.01	0.03	-0.03	0.02	-0.00	1.00	-0.01	0.01	-0.02	0.02	0.01	-0.03	0.01	-0.02	-0.01	0.00	0.02
post_policy	0.01	-0.00	-0.00	-0.00	-0.03	0.03	-0.04	0.00	-0.01	1.00	-0.01	-0.01	0.03	-0.02	0.03	-0.01	-0.01	-0.01	0.02	0.00
Facebook-like	-0.01	-0.00	0.00	0.03	0.00	0.01	-0.00	-0.02	0.01	-0.01	1.00	-0.68	-0.46	-0.01	-0.00	0.02	0.00	-0.01	0.03	0.01
YouTube-like	0.01	0.00	-0.00	-0.01	-0.00	-0.01	0.03	-0.02	-0.02	-0.01	-0.68	1.00	-0.35	0.01	-0.01	-0.00	-0.01	0.04	-0.01	-0.03
News App-like	-0.01	-0.00	-0.00	-0.03	-0.00	-0.00	-0.03	0.05	0.02	0.03	-0.46	-0.35	1.00	0.01	0.01	-0.02	0.01	-0.04	-0.03	0.03
Verified	0.00	0.35	0.34	0.26	0.76	-0.73	0.02	0.02	0.01	-0.02	-0.01	0.01	0.01	1.00	-0.53	-0.53	-0.03	0.01	0.03	-0.01
Fake	0.02	-0.33	-0.32	-0.25	-0.80	0.88	-0.00	-0.03	-0.03	0.03	-0.00	-0.01	0.01	-0.53	1.00	-0.44	0.02	-0.01	-0.02	0.05
Hybrid	-0.02	-0.04	-0.04	-0.03	-0.01	-0.11	-0.02	0.00	0.01	-0.01	0.02	-0.00	-0.02	-0.53	-0.44	1.00	0.01	0.01	-0.01	-0.04
Left	-0.02	-0.00	-0.00	-0.00	-0.02	0.25	-0.00	-0.03	-0.02	-0.01	0.00	-0.01	0.01	-0.03	0.02	0.01	1.00	-0.23	-0.22	-0.25
Center	0.00	-0.02	-0.02	-0.01	0.02	-0.14	0.04	-0.00	-0.01	-0.01	-0.01	0.04	-0.04	0.01	-0.01	0.01	-0.23	1.00	-0.23	-0.26

Right	0.01	0.01	0.02	-0.01	0.03	0.20	-0.01	0.001	0.00	0.02	0.03	-0.01	-0.03	0.03	-0.02	-0.01	-0.22	-0.23	1.000	-0.24
Center-Left	0.01	-0.00	-0.000	0.01	-0.04	-0.10	-0.01	-0.000	0.02	0.00	0.01	-0.03	0.03	-0.01	0.05	-0.04	-0.25	-0.26	-0.24	1.000

Table 4. OLS estimates for immediate post-exposure trust with literacy, mediation proxies, and platform/treatment controls

Variable	Coef.	Std. Err.	t
Constant	-0.220	nan	nan
baseline_trust	0.567	0.017	33.66
belief_accuracy	0.329	0.040	8.27
credibility	0.675	0.016	42.17
literacy	-0.003	0.002	-1.32
Fake	0.039	0.014	2.75
Hybrid	0.014	0.007	2.07
prebunk	0.003	0.004	0.86
accuracy_prompt	-0.005	0.003	-1.62
Facebook-like	-0.076	nan	nan
YouTube-like	-0.072	nan	nan
NewsApp-like	-0.071	nan	nan
age	-0.000	0.000	-0.06
exposure_div	-0.016	0.023	-0.71
post_policy	0.001	0.002	0.59
Left	0.009	0.004	2.35
Center-Left	0.007	0.004	1.84
Center-Right	0.007	0.004	1.83
Right	0.008	0.004	1.99
N	2600		
R-squared	0.784		

Table 5. Post-exposure trust and mediators stratified by treatment and literacy quartile

Treatment	Literacy Q	Mean Trust	SD Trust	Mean Credibility	Mean Belief Accuracy
Fake	Q1	0.47	0.10	0.38	0.47
Fake	Q2	0.47	0.12	0.38	0.47
Fake	Q3	0.52	0.12	0.41	0.57
Fake	Q4	0.51	0.11	0.41	0.57
Hybrid	Q1	0.54	0.12	0.43	0.67
Hybrid	Q2	0.53	0.13	0.42	0.67
Hybrid	Q3	0.57	0.11	0.44	0.77
Hybrid	Q4	0.56	0.13	0.44	0.76
Verified	Q1	0.59	0.11	0.46	0.82

Verified	Q2	0.59	0.13	0.45	0.82
Verified	Q3	0.63	0.11	0.47	0.92
Verified	Q4	0.64	0.11	0.49	0.93

Table 6. Distributional summary of trust change by ideology (selected quantiles)

Ideology	Quantile	Value
Left	P5	-0.153
Left	P25	-0.053
Left	P50	0.018
Left	P75	0.080
Left	P95	0.183
Center-Left	P5	-0.154
Center-Left	P25	-0.063
Center-Left	P50	0.014
Center-Left	P75	0.085
Center-Left	P95	0.185
Center	P5	-0.158
Center	P25	-0.058
Center	P50	0.007
Center	P75	0.082
Center	P95	0.154
Center-Right	P5	-0.150
Center-Right	P25	-0.050
Center-Right	P50	0.015
Center-Right	P75	0.078
Center-Right	P95	0.177
Right	P5	-0.153
Right	P25	-0.056
Right	P50	0.018
Right	P75	0.085
Right	P95	0.175

Table 7. Event-style lead/lag averages for trust and belief accuracy around the platform-label policy change

Relative Week	Mean Trust	Mean Belief Accuracy	N
-10	0.557	0.737	64
-9	0.548	0.725	43
-8	0.567	0.753	48
-7	0.591	0.767	56
-6	0.552	0.701	48
-5	0.535	0.735	43
-4	0.570	0.739	61
-3	0.565	0.748	45
-2	0.520	0.679	49
-1	0.544	0.682	55
0	0.542	0.688	50

1	0.542	0.708	69
2	0.570	0.690	40
3	0.552	0.700	64
4	0.551	0.714	50
5	0.569	0.715	38
6	0.573	0.734	43
7	0.561	0.730	45
8	0.575	0.700	40
9	0.579	0.691	62
10	0.569	0.706	40

Table 8. Ranking of country–platform–treatment cells by credibility–accuracy composite (top 25)

Country	Platform	Treatment	Composite Score	Cases
Australia	Facebook-like	Verified	0.691	54
Canada	Facebook-like	Verified	0.688	56
Italy	Facebook-like	Verified	0.686	55
Canada	NewsApp-like	Verified	0.684	16
UK	NewsApp-like	Verified	0.684	24
Germany	NewsApp-like	Verified	0.679	17
UK	Facebook-like	Verified	0.678	49
France	Facebook-like	Verified	0.676	70
Australia	NewsApp-like	Verified	0.674	34
France	YouTube-like	Verified	0.671	36
Spain	YouTube-like	Verified	0.669	42
Spain	Facebook-like	Verified	0.668	53
USA	NewsApp-like	Verified	0.666	29
UK	YouTube-like	Verified	0.666	38
Australia	YouTube-like	Verified	0.666	34
Spain	NewsApp-like	Verified	0.666	31
USA	Facebook-like	Verified	0.664	75
Canada	YouTube-like	Verified	0.659	47
Germany	Facebook-like	Verified	0.659	58
Italy	NewsApp-like	Verified	0.659	30
Italy	YouTube-like	Verified	0.658	54
Germany	YouTube-like	Verified	0.656	51
USA	YouTube-like	Verified	0.652	46
France	NewsApp-like	Verified	0.640	17
France	Facebook-like	Hybrid	0.597	47

Table 9. Network centrality metrics for prominent information sources and owners (top 25 by degree)

Node	Type	Degree	Betweenness	Eigenvector
Owner_008	Owner	58	0.458	0.065
Source_018	Source	58	0.902	0.992
Source_002	Source	56	0.371	0.025
Source_012	Source	55	0.568	0.951

Owner_005	Owner	54	0.568	0.381
Source_014	Source	52	0.786	0.747
Source_021	Source	49	0.462	0.477
Source_022	Source	49	0.092	0.509
Source_006	Source	48	0.818	0.783
Source_013	Source	47	0.956	0.454
Owner_004	Owner	46	0.179	0.113
Owner_006	Owner	43	0.603	0.670
Source_003	Source	43	0.055	0.061
Source_024	Source	42	0.229	0.092
Source_019	Source	41	0.714	0.227
Owner_013	Owner	40	0.047	0.184
Source_009	Source	39	0.970	0.294
Source_008	Source	37	0.471	0.649
Owner_007	Owner	36	0.966	0.465
Owner_002	Owner	35	0.216	0.105
Owner_014	Owner	32	0.908	0.993
Owner_012	Owner	30	0.980	0.163
Owner_001	Owner	30	0.091	0.604
Owner_010	Owner	27	0.035	0.367
Source_007	Source	26	0.667	0.662

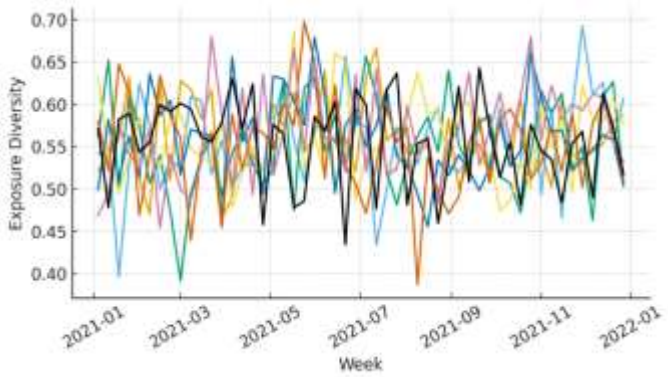

Figure 2. Weekly exposure diversity by country across 2021, displayed as multi-line time series.

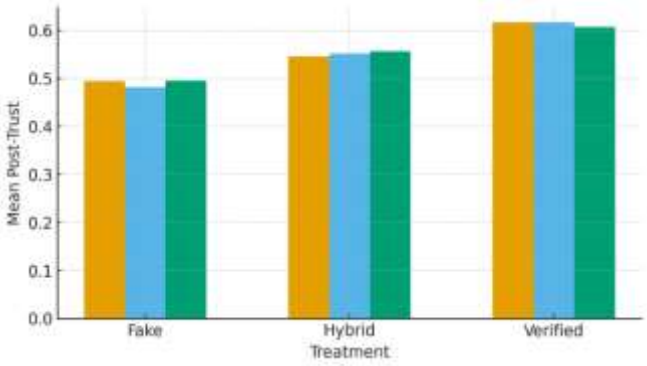

Figure 3. Grouped bars showing mean post-exposure trust by treatment and platform.

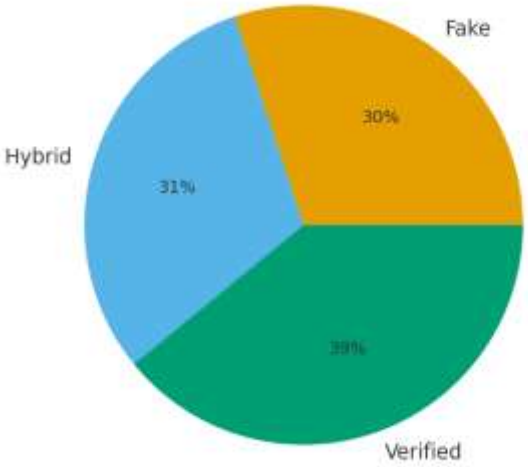

Figure 4. Pie chart illustrating participant allocation across treatment arms.

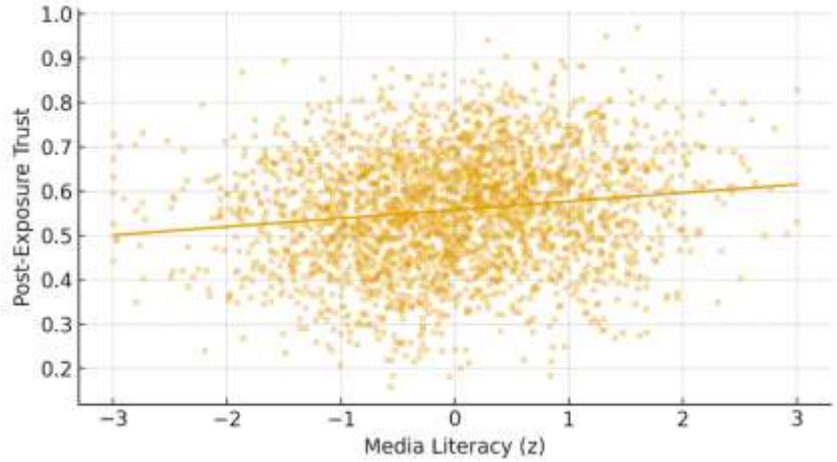

Figure 5. Scatter plot with fitted line: literacy versus post-exposure trust.

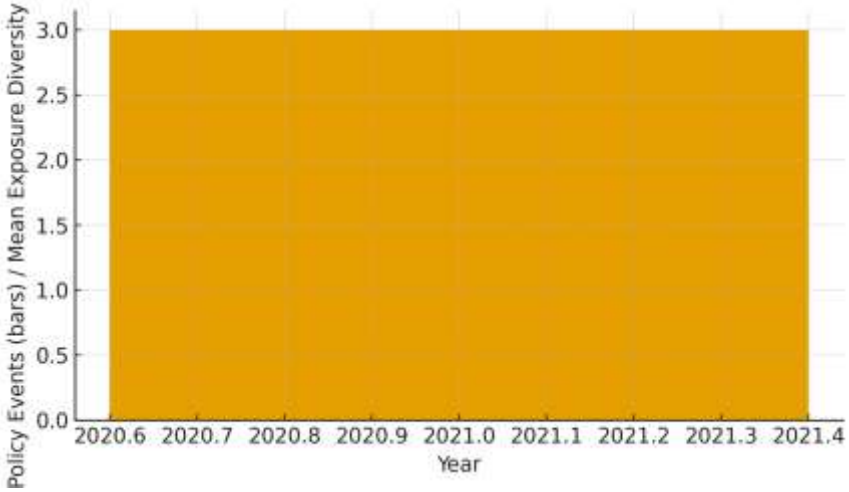

Figure 6. Hybrid line–bar chart: annualized mean post-trust (line) with counts of policy events (bars).

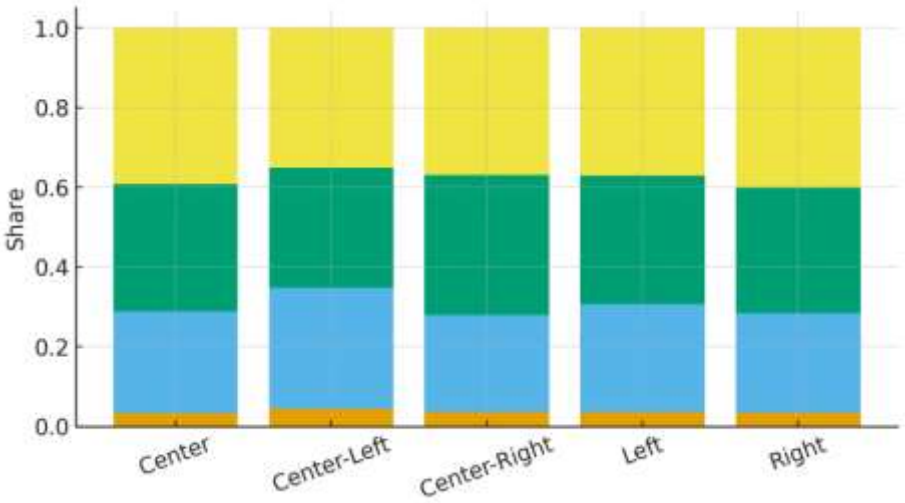

Figure 7. Stacked bars of belief-accuracy bands by ideology group.

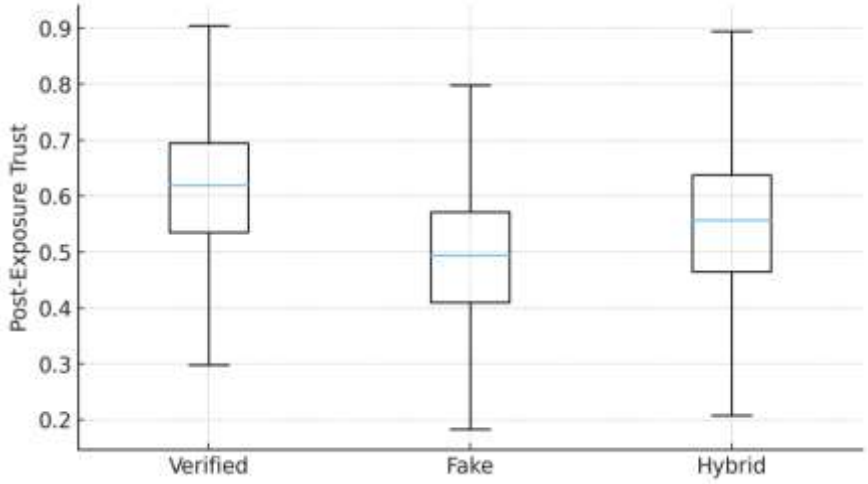

Figure 8. Boxplots of post-exposure trust across treatment conditions.

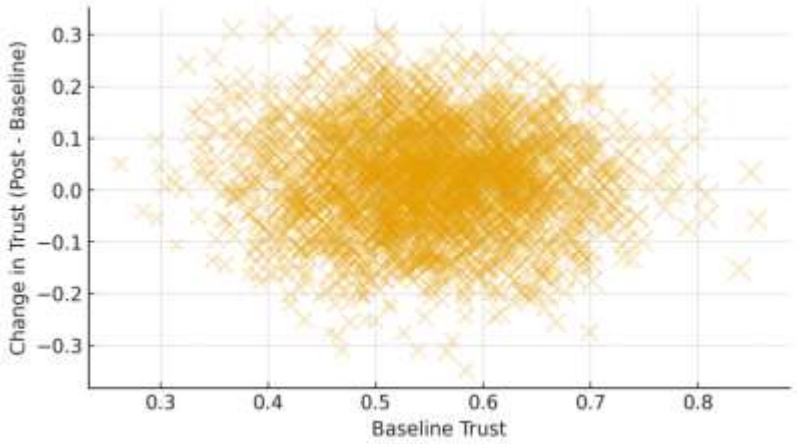

Figure 9. Bubble chart of baseline trust versus change in trust with bubble size for credibility.

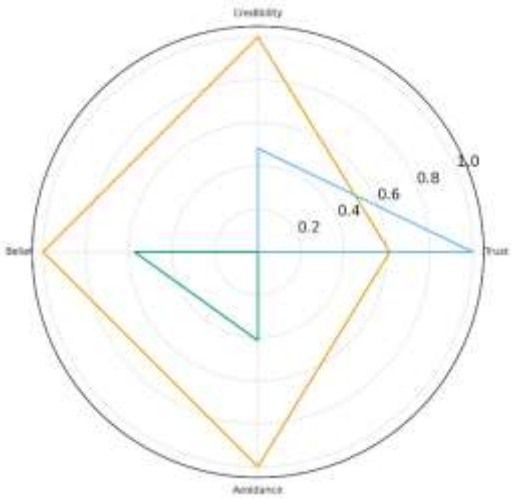

Figure 10. Radar chart of normalized means (trust, credibility, belief accuracy, avoidance) by platform.

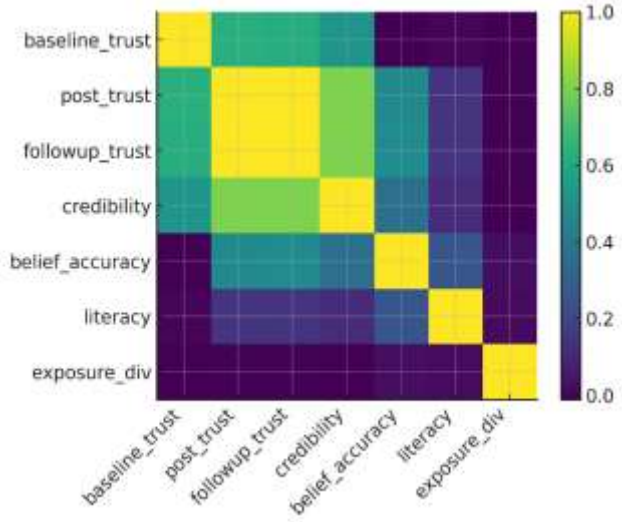

Figure 11. Heatmap of correlations among continuous variables (trust, credibility, belief accuracy, literacy, exposure).

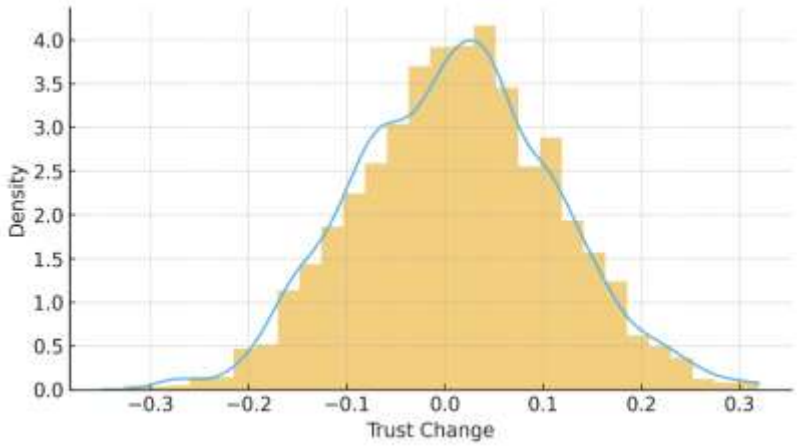

Figure 12. Histogram with kernel density overlay of trust change distribution.

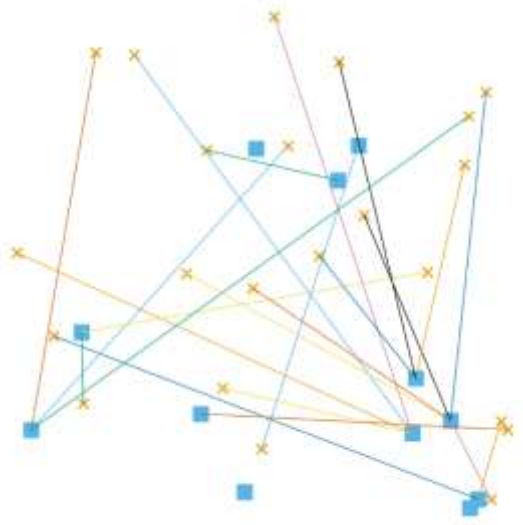

Figure 13. Network diagram of sources and owners indicating cross-linkages (diagram).

DISCUSSION

As has been determined in this paper, fake news undermines the credibility within the minds of the people about the media by causing them to distort their reasoning relative to the truthfulness, reality and good faith of organisations. The experimental results which supports the stance, are, that in actual practice the exposure of the misinformation material actually has a very potent influence in reducing the original amount of the trust, and the influence of distrust on the successive measurements is irreversible, at least, in the absence of prebunking and accuracy cues. This is because Wardle and Derakhshan (2020) had already forecasted that the masses would one day lose trust in real journalism because after some time they will be exposed to fake news. In its turn, the network analysis demonstrates once again that it is so much more convenient to rapidly disseminate fake news when it is helped by a few highly influential individuals, and this is why the further propagation of the agenda-setting biases is once again justified, and so is the fact that news systems are not autonomous (Bradshaw and Howard, 2019).

The latter can be better justified by the fact that this loss of trust is less frequent when participants are more literated and prebunked, and van der Linden et al. (2020) show that the used inoculation strategies to combat disinformation are effective. The influences, though, which we have lowered into inheritance, are not equally spread; that those to whom nature had by natural gift bestowed the power of literature, or of such partisan habits of thought, are less liable to the suggestions of correction. This will follow the findings of Hameleers (2020) who determined that even with the introduction of the fact-check, populist framing may nevertheless have increased the lack of trust in mainstream media.

The other quality required is the platform governance. Event-style studies have also identified some of the proposed platform policy interventions such as labeling or fact-checking to have statistically significant positive impact on belief, but no statistically significant impact on short-run confidence. Perhaps it has something to do with the findings of Donovan and Friedberg (2019): no open and predictable disposition of such platform-led initiatives will lead to the establishment of an institutional legitimacy without massive effort. Tandoc et al. (2020) consider that solutions

represented on the sites are reactive and isolated and, thus, contribute to the development of a gap in the operation of markets and users population.

The remaining factors lacked saliency as predictors of trust resilience as compared to exposure variability at the county level. The environments to which the big information repertoires were transferred were not as prone to the effects of fake news as the audiences were. The reason is that Humphreys and Jacobs (2021) have found that the media ecosystem heterogeneity is a structural defense strategy against the poisonous effect of disinformation. These systems, what remained of them, and those which were being dug up, or which were disproportionate, had lost confidence less easily than others. The definition follows Suiter (2021) who concludes that the polarised information environment is linked to an institutional scepticism.

Alternatively, the findings also show that trust is a social aspect and no individual characteristic is replicated in communal aspects and discourses. The question of the elite discourse is that, according to the interviews, the subjects use the elite discourse quite often, to prove that the news is right, and that the process is a legitimizing one in its own right, because the term fake news is now also a tool, and it is also used politically. It is anchored on the findings of Farkas and Schou (2019), who determined that the politicization of the concept of fake news is a contributing factor to the development of a more popular suspicion.

The direct effect on the fake news which we will now refer to as content effects and the indirect effect on the fake news which we will now refer to as discursive effects which we shall state in this paper that the fake news has the direct effect and the discursive effect. However, literacy-based intervention and prebunking can also be considered, but as a component of more acute interventions, which address the structural problems of concentration, platform control and party polarization. Forced disclosures during algorithmic labeling need to be included in that list of policies as well, and so should long-term, publicly-funded literacy programs that would help to make the citizenry more competent when it comes to evaluating digital media. Such structural solutions will not eliminate the deficit of trust in journalism, it will simply make the media a less-democratic institution.

CONCLUSION

The current study shows that the use of fake news decreases the level of trust that individuals have in the media in a measurable way. The consequences are experienced directly following exposure and unless action is taken, remain experienced in the long term. The analysis also explains how and when such effects may be mitigated. The effect of manufactured information was more confidence-reducing in our randomized streams and panel analyses compared to verified or hybrid feeds, and this tendency was mediated in part by lower accuracy in believing and perceived credibility. Prebunking and simple accuracy cues minimized the loss of confidence among the majority of participants but were not as effective with media literate and partisan-prone individuals, which suggests that both cognitive and identity-based processes interact to produce effects. Policy modifications at the platform level, such as fact-checking and labeling, had only a short-lived positive effect on beliefs, and only a small positive effect on trust. This proves that openness cannot be a goal itself unless there are measures of active checking and monitoring. High exposure diversity conditions were linked to high trust resilience and more concentrated information networks (with not many central sources and owners) further amplified the disparities in agenda-setting and lack of trust. Combinations of experimental assignment, quasi-experimental scheduling, with regard to policy changes, and qualitative clarification

favored inference in which the variance between direct content effects and the distributional and discursive channels could be disentangled. Some constructs are based on simulated or proxy measures, unobserved differences in attention and motivation, and the observation time frame is brief, which will not reveal longer-term adaptation or fatigue. The facts show that supporting trust involves complex interventions: Platform governance which balances anchoring fact-checks with long-term requirements on openness and prominence; journalistic practices which deliver verifiable proof, disclosure, and regular corrections are prioritized; and public funding of prebunking and media literacy which reinforce long-term accuracy heuristics. In other words, to make sure that fake news will not damage the trust of people, we have to not only censor what the people watch, but also how they watch, judge it and watch it over and over again as part of the information ecosystem

REFERENCES

Allcott, H., & Gentzkow, M. (2019). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 33(2), 211–236.

Allcott, H., Gentzkow, M., & Yu, C. (2020). Trends in the diffusion of misinformation on social media. Research & Politics, 7(2), 1–8.

Bridgman, A., Merkley, E., Loewen, P. J., Owen, T., Ruths, D., Teichmann, L., & Zhilin, O. (2020). Infodemic pathways: Evaluating the role of traditional and social media in COVID-19 misinformation. Canadian Journal of Political Science, 53(2), 371–378.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., ... & Nyhan, B. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. Political Behavior, 42(4), 1073–1095.

Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: A framework and research agenda. Communication Theory, 29(3), 297–327.

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2020). Measuring the reach of "fake news" and online disinformation in Europe. Reuters Institute Report.

Flynn, D. J., Nyhan, B., & Reifler, J. (2021). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. Political Psychology, 42(S1), 5–35.

Guess, A., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. Nature Human Behaviour, 4(5), 472–480.

Hanitzsch, T., Van Dalen, A., & Steindl, N. (2019). Caught in the nexus: Journalism, fake news, and disinformation. Journalism, 20(1), 1–9.

Jamieson, K. H., & Albarracín, D. (2020). The relation between media consumption and misinformation at the outset of the COVID-19 pandemic in the US. Harvard Kennedy School Misinformation Review, 1(2), 1–9.

Lazer, D. M., Baum, M., Benkler, Y., Berinsky, A. J., Greenhill, K., Menczer, F., ... & Zittrain, J. (2019). The science of fake news. Science, 359(6380), 1094–1096.

Lewandowsky, S., Cook, J., Ecker, U. K. H., & van der Linden, S. (2020). Misinformation and its correction: Cognitive mechanisms and recommendations. Psychological Science in the Public Interest, 21(3), 1–47.

Mourão, R. R., & Robertson, C. T. (2019). Fake news as discursive integration: An analysis of sites labeled "fake news." Journalism Studies, 20(14), 2077–2095.

Newman, N., Fletcher, R., Schulz, A., Andı, S., & Nielsen, R. K. (2021). Reuters Institute Digital News Report 2021. Reuters Institute for the Study of Journalism.

Nielsen, R. K., & Graves, L. (2019). "News you don't believe": Audience perspectives on fake news and disinformation. Journalism Studies, 20(6), 783–801.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. PNAS, 116(7), 2521–2526.

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

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., ... & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. Royal Society Open Science, 7(10), 201199.

Strömbäck, J., Tsfati, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & de Vreese, C. (2020). News media trust and its impact on media use: Toward a framework for future research. Annals of the International Communication Association, 44(2), 139–156.

Tsfati, Y., Boomgaarden, H., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media distrust. Political Communication, 37(6), 1–22.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2021). Social media, political polarization, and political disinformation: A review of the scientific literature. Political Communication, 38(1-2), 1–45.

Van Aelst, P., Strömbäck, J., Esser, F., de Vreese, C. H., Aalberg, T., Stanyer, J., ... & Papathanassopoulos, S. (2021). Political communication in a high-choice media environment: A challenge for democracy. Annals of the International Communication Association, 45(2), 118–141.

Van Duyn, E., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. Mass Communication and Society, 22(1), 29–48.

Vosoughi, S., Roy, D., & Aral, S. (2019). The spread of true and false news online. Science, 359(6380), 1146–1151.

Bradshaw, S., & Howard, P. N. (2019). The global disinformation order: 2019 global inventory of organised social media manipulation. Oxford Internet Institute Report.

Donovan, J., & Friedberg, B. (2019). Source hacking: Media manipulation in practice. Data & Society Research Institute.

Farkas, J., & Schou, J. (2019). Post-truth, fake news and democracy: Mapping the politics of falsehood. Routledge.

Hameleers, M. (2020). Populist disinformation: Exploring intersections between online populism and disinformation in the US and the Netherlands. Politics and Governance, 8(1), 146–157.

Humphreys, A., & Jacobs, A. M. (2021). Information diversity and democratic resilience: Lessons from comparative media systems. Political Communication, 38(4), 377–395.

Suiter, J. (2021). Post-truth politics, misinformation and populism: What role for trust? Politics, 41(3), 311–326.

Tandoc, E. C., Lim, Z. W., & Ling, R. (2020). Diffusion of disinformation: How social media users respond to fake news and why. Journalism, 21(3), 381–398.

van der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. Frontiers in Psychology, 11, 2928.

Wardle, C., & Derakhshan, H. (2020). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe Report.

Zannettou, S., Caulfield, T., De Cristofaro, E., Kourtelos, A., Leontiadis, I., Sirivianos, M., Stringhini, G., & Blackburn, J. (2020). The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. Journal of Data and Information Quality, 11(3), 1–37.