

RESEARCH ARTICLE  

# Inoculation Reduces Misinformation: Experimental Evidence from Multidimensional Interventions in Brazil

Frederico Batista Pereira<sup>1</sup>, Natália S. Bueno<sup>2</sup>, Felipe Nunes<sup>3</sup> and Nara Pavão<sup>4</sup> 

<sup>1</sup>University of North Carolina at Charlotte, Charlotte, NC 28223, USA, <sup>2</sup>Emory University, Atlanta, GA 30322, USA, <sup>3</sup>Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil and <sup>4</sup>Universidade Federal de Pernambuco, Recife, PE 50670-901, Brazil

**Corresponding author:** Natália S. Bueno; Email: [natalia.bueno@emory.edu](mailto:natalia.bueno@emory.edu)

## Abstract

Misinformation is widely seen as a threat to democracy that should be promptly addressed by scholars, journalists, and policymakers. However, some of the debated solutions are either controversial (internet platform regulation) or may be difficult and costly to implement in many settings (fact-checking corrections). This study investigates the effectiveness of preemptive interventions, a type of solution that has received considerably less attention in this debate. Studies show that interventions through awareness and media literacy campaigns can inoculate citizens against misinformation, but these interventions are restricted to a few contexts and settings. Our paper uses two field experiments, one of which was conducted in partnership with Brazil's main newspaper, to investigate the effectiveness of multidimensional interventions against misinformation in São Paulo. The findings show that preemptive interventions can indeed reduce rumor acceptance and provide insights into the strategies to combat misinformation in democracies.

**Keywords:** fake news; misinformation; field experiment; inoculation

Growing concern about online misinformation has given rise to debates about possible interventions to mitigate the spread of fake news. Some common solutions considered by scholars, journalists, and policymakers, such as fact-checking corrections and internet platform regulation, seek to reduce the dissemination of misinformation in the news market. While some argue that these solutions have limited effectiveness or tend to be controversial and unfeasible in the short run (McIntery 2018), less attention has been given to solutions that target citizens' propensity to believe misinformation without providing corrective information or limiting the

  This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the [Data Availability Statement](#).

© The Author(s), 2023. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

spread of information online. In this paper, we investigate a specific type of intervention against misinformation in the news market. Preventive interventions, such as awareness and media literacy campaigns, seek to motivate and train individuals to critically engage with news content and distinguish between accurate and inaccurate information (Bulger and Davidson 2018). Since such interventions do not attempt to correct specific rumors, which are often tied to political groups and identities, they may be effective at inoculating citizens against misinformation. Despite a burgeoning scholarship on experimental assessments of their effects against misinformation, field experimental interventions are not too common (for exceptions, see Badrinathan 2021; Guess *et al.* 2020, among others).

We implement two experiments embedded in online survey panels to test the effectiveness of a campaign against misinformation. The first was carried out during the 2020 mayoral elections in São Paulo, Brazil. The intervention was designed through a partnership with the largest Brazilian newspaper, *Folha de São Paulo*, which provided free subscriptions and email content to be used in the treatment stimuli. The second study, fielded in January 2022, replicated features of the first study and included some additional elements. Our findings show that the interventions decreased rumor acceptance, thus providing further evidence that awareness and media literacy campaigns can reduce people's belief in misinformation. The results are particularly striking in the context of Brazil, where misinformation is widespread and evidence of fact-checking effectiveness is mixed (Batista Pereira *et al.* 2022; Carey *et al.* 2020; Machado *et al.* 2019; Porter *et al.* 2023).

## Theory

Preemptive media literacy and awareness interventions seek to increase individuals' motivation and skills to critically engage with media content (Bulger and Davidson 2018). While these interventions are most commonly carried out in educational settings, the growing availability of social media to larger shares of the public has led to a number of initiatives attempting to reach broader audiences (Lee 2018). Moreover, organizations such as Whatsapp and Facebook have recently launched large-scale cross-national campaigns to educate their users on misinformation (Guess *et al.* 2020).

Much like a vaccine against a disease, preemptive messages aim at protecting individuals from being persuaded by misinformation (McGuire 1961). Additionally, inoculation can operate through motivational or refutational mechanisms (Banas and Miller 2013). While motivational interventions warn individuals about their potential vulnerability to attitude change in the face of persuasion, refutational preventive interventions seek to provide informational content that undermines the effect of future persuasion attempts.

The theory underlying this type of intervention is consistent with classic dual-process models of attitude change and persuasion (Eagly and Chaiken 1993; Petty and Cacioppo 1996). According to this perspective, human cognition can be divided into intuitive/automatic and deliberate/analytical processes of stimuli and information. As Pennycook and Rand (2019) show, rumor acceptance is largely associated with intuitive/automatic processes by which individuals process

information. In this sense, preemptive interventions can reduce rumor acceptance by nudging individuals to rely on more deliberate/analytical forms of cognitive processing when they encounter new information. Consequently, preemptive interventions can affect individuals' willingness to engage in effortful thinking to avoid accepting false information, especially when it reinforces prior beliefs (Pennycook and Rand 2019). Those interventions, which can involve a variety of literacy skills to evaluate media content and online information, can promote the skill sets required to engage with media content and resist content containing misinformation (Jones-Jang et al. 2021).

Extant scholarship provides experimental evidence of the effects of media literacy interventions on different skills related to engagement with media content, such as knowledge and self-efficacy (Jeong et al. 2012). Among observational studies, the evidence suggests that media literacy is associated with lower probability of accepting rumors and conspiracy theories (Jones-Jang et al. 2021). Similar results are observed in experimental assessments focusing on misinformation related to climate change (Cook et al. 2017; Lutzke et al. 2019) and genetically modified foods (Tully et al. 2020).

Recent scholarship using experimental approaches—restricted to a few countries and settings—finds that preemptive interventions tend to reduce demand for misinformation, but findings are somewhat mixed. On the one hand, Banas and Miller (2013) find that interventions reduce acceptance of 9/11 conspiracy theories among college students. Similarly, Roozenbeek and van der Linden (2019) use an online game as an intervention and find reduction in rates of rumor acceptance among players across different countries. Guess et al. (2020) show that a Facebook media literacy initiative against misinformation reduces rumor acceptance in the USA and India.

On the other hand, Hameleers (2022) finds statistically significant but weak effects of a media literacy intervention on reducing the perceived accuracy of false news in Netherlands and the USA. Clayton et al. (2020) find that a general “awareness warning” in a US sample reduces rumor acceptance, but more strongly when paired with fact-checking tags. Finally, Badrinathan (2021) does not find evidence that an educative intervention on information processing was sufficient to counter misinformation in India.

Finally, assessing the effectiveness of preemptive messages carries special normative relevance in the context of this study. In Brazil, as in several developing nations, the spread of misinformation occurs primarily through instant messaging cell phone applications, such as Whatsapp (Rossini et al. 2021), making platform regulation and corrections by automated algorithms less viable large-scale solutions in that context.

## Hypotheses

We assess the effectiveness of preemptive interventions against misinformation using two field experiments in Brazil. We expect the intervention described in the next section to reduce rumor acceptance by providing individuals with the opportunity, motivation, and ability to reject inaccurate information. We also

expect the intervention to have no effect on subjects' propensity to reject factually accurate information.<sup>1</sup>

**Hypothesis 1 (H1).** *The preemptive intervention will reduce rumor acceptance among subjects in the treatment group relative to subjects in the control group.*

**Hypothesis 2 (H2).** *The preemptive intervention will not reduce acceptance of correct information among subjects in the treatment group relative to subjects in the control group.*

## Study 1

Study 1 uses a two-wave online survey combined with an experimental intervention conducted during the 2020 mayoral elections in São Paulo, Brazil.<sup>2</sup> The intervention was designed in partnership with *Folha de São Paulo*, a newspaper founded in 1921 with the largest circulation and online presence in Brazil. Data collection was performed by *Quaest Consultoria & Pesquisa*, which has an online panel of approximately 150,000 people in more than 2,000 cities in Brazil. The first wave of the survey included 1,000 respondents and was conducted between November 19 and 24, a few days before the second round of the local elections (November 29). The second wave took place between December 8 and 16 and we were able to re-interview 731 respondents from the first wave.

After completing the first wave questionnaire, respondents were randomly assigned, via simple randomization, to one of two conditions. After the first wave of the survey and before the second, the treatment group ( $n = 575$ ) received the main experimental stimuli comprised of a multidimensional intervention seeking to reduce rumor acceptance. As the first component of the intervention, the treatment group received at the end of the survey a voucher for a free 3-month subscription to *Folha de São Paulo*. This component of the intervention offered participants the opportunity to learn and distinguish between factually correct and incorrect information, since the newspaper provides professional news coverage and fact checks the main rumors circulated in Brazil.<sup>3</sup> A week after the first survey wave (December 1), the treatment group also received an email from the survey company containing a message about the spread of fake news during the election. The message included a link to a news piece from *Folha de São Paulo* that described the newspaper's fact-checking tools available for respondents. The email also included a list of eight steps for spotting fake news. The email sought to foster both motivation and ability to reject misinformation, since the warning message encouraged respondents to be careful with online content and the 8-step recommendations

---

<sup>1</sup>Our experiments were not pre-registered.

<sup>2</sup>Both studies 1 and 2 were deemed exempt by the IRBs of the University of North Carolina at Charlotte, Emory University, and Getúlio Vargas Foundation.

<sup>3</sup>Since the rumors included in the surveys circulated prior to its design, they were not fact-checked by the website between the two waves, but had already been publicly debunked by many fact-checking agencies.

provided skills to evaluate online content.<sup>4</sup> The control group ( $n = 425$ ) did not receive any stimuli related to the campaign.

The survey measured rumor acceptance by presenting four rumors that contained factually incorrect information (according to multiple professional fact-checkers) and then asking whether respondents believed they were true or false.<sup>5</sup> We used two different sets of four rumors in each wave of the survey, with one rumor repeating between waves.<sup>6</sup> We assess whether repeated exposure influences treatment effects by analyzing the effects on the gain scores between the first and second rounds for the repeated rumor.<sup>7</sup> In addition to the new set of rumors and a repeated rumor, the second wave of Study 1 included questions on trust in media and attitudes toward fact-checking.<sup>8</sup>

Since subjects assigned to treatment could choose to ignore the voucher offered by the newspaper and the information sent via e-mail (Gerber and Green 2012, 131), we had cases of noncompliance. Out of 575 subjects in the group assigned to treatment, 66 activated the newspaper voucher and completed their registration in the newspaper's website, and 272 subjects opened the follow-up email.<sup>9</sup> In total, 304 subjects complied by either opening the email or accessing the newspaper's website.<sup>10</sup> We present both intent-to-treat (ITT) estimates and complier average causal effect (CACE) estimates to measure the impact of the intervention in the group of all study subjects and in the subgroup that complies with the assigned treatment, respectively.<sup>11</sup>

We find that treatment assignment is statistically associated with a few pre-treatment covariates, including rumor acceptance. We performed three joint statistically significant tests ( $F$ , Chi-Square, and Hotelling's  $T$ ), one of them ( $F$ -test) was statistically significant at levels lower than 0.05 and all three were equal or lower than levels of 0.10.<sup>12</sup> We further examined the randomization algorithm used by the polling company and found no systematic error in the randomization procedure.<sup>13</sup> We show the main results unadjusted for covariates, but results do not change when we control for pre-treatment variables.<sup>14</sup>

<sup>4</sup>See Section 1 of the supplemental appendix for the email sent to respondents.

<sup>5</sup>See Section 2 and Section 3 of the supplemental appendix for full wording (in English and Portuguese, respectively) of our outcome questions.

<sup>6</sup>Presenting the same rumors twice can be concerning because repeated exposure may increase rumor acceptance (Fazio et al. 2019; Pennycook et al. 2018) and generate demand effects. To minimize these ethical and methodological concerns, in the two studies we repeated just a small set of the rumors used in wave 1 (1 repetition in study 1 and 3 repetitions in study 2).

<sup>7</sup>See Section 2 and Section 3 of the supplemental appendix for instrumentation (in English and Portuguese, respectively) containing the rumors used in each round of the survey.

<sup>8</sup>These attitudes represent individuals' assessments of how helpful fact-checking is, the frequency with which they use fact-checking during elections, and how reliable fact-checking agencies are in Brazil.

<sup>9</sup>A total of 203 subjects activated the voucher, but only 66 filled out the required information to receive the subscription. The information on voucher activation was provided by *Folha de São Paulo*. The subjects who opened the email were tracked by the survey company.

<sup>10</sup>Of those 304 subjects that complied in at least one way, 222 participated in the second round of the survey (55% of the treatment group in round 2).

<sup>11</sup>See Section 4 of supplemental appendix for descriptive statistics.

<sup>12</sup>See Section 5 of the supplemental appendix.

<sup>13</sup>See Section 6 of the supplemental appendix for a report on that issue.

<sup>14</sup>See Section 7 of the supplemental appendix for models adjusted for pre-treatment covariates.

We were able to re-interview about 73% of our first-round sample. However, we found evidence of differential panel attrition, with respondents from the group assigned to treatment being more likely to drop out in the second wave of the survey.<sup>15</sup> To account for potential biases from differential attrition, we also assess the results using different bounding approaches (Gerber and Green 2012, 226).<sup>16</sup>

We present our results for two main dependent variables.<sup>17</sup> The first is an additive scale of rumor acceptance for the four rumors presented in the second wave of the survey (each coded as 0 for rejection and 1 for acceptance). The variable is re-scaled from 0 to 1 and indicates the extent to which subjects accept false rumors as true. The second dependent variable indicates the change in response to the repeated rumor, taking values  $-1$  (believed in wave 1 but not in wave 2),  $0$  (did not change response between waves), and  $1$  (did not believe in wave 1 and believed in wave 2).

We assess the effect of the study's intended treatment assignments, or intent-to-treat (ITT) effects, by using ordinary least squares (OLS) models (with robust standard errors). To estimate the complier average treatment effects (CACE), we use two-stage least squares (2SLS) regressions with robust standard errors for three different measures of compliance instrumented in the first-stage regression, while having the treatment assignment as the instrument in the second-stage regression (Angrist *et al.* 1996; Gerber and Green 2012). The three measures of compliance refer to whether subjects accessed (activated the voucher and registered for) the newspaper's website, whether subjects opened the follow-up email, and whether they complied by either accessing the newspaper or opening the email. The results are presented in Table 1.

The intent-to-treat estimates show that the intervention reduced on average 0.12 points in the additive scale of rumor acceptance (column 1)<sup>18</sup> and 0.24 points in acceptance of the repeated rumor (column 5).<sup>19</sup> The complier average causal effects also show the expected negative effects for the two outcomes. All of our measures of compliance, by whether subjects used the newspaper subscription (columns 2 and 6), opened the follow-up email (columns 3 and 7), or both combined (columns 4 and 8), indicate the intervention decreased the levels of rumor acceptance among compliers. Overall, our results are consistent with the idea the intervention makes subjects more likely to reject false rumors.

In regards to moderation, the treatment effects (ITT) do not show consistent patterns for both the additive scale and the repeated rumor with respect to pre-treatment levels of political knowledge, trust in the media, political interest, and support for the President. Hence, the intervention was not significantly less effective among subgroups that could be theoretically more prone to reject it.<sup>20</sup> We also find that the intervention has a significant effect on rumors in favor of President Bolsonaro and not on those against him. However, the reduced number of rumors used per type (2) suggests that the results should be taken with caution, since they could be dependent on specific rumor characteristics. Last, the intervention did not

<sup>15</sup>See Section 8 of the supplemental appendix for an analysis of panel attrition.

<sup>16</sup>Our bounding procedures show that the results hold under several different assumptions about differential attrition. See Section 9 of supplemental appendix for results.

<sup>17</sup>See footnote 6 for explanation on why we use two dependent variables.

<sup>18</sup>Cohen's  $d$  for additive scale is  $-0.50$  [ $-0.65$ ;  $-0.35$ ].

<sup>19</sup>Cohen's  $d$  for gain in repeated rumor is  $-0.53$  [ $-0.68$ ;  $-0.39$ ].

<sup>20</sup>See Section 10 for analyses of heterogeneous effects.

**Table 1.** Intent-to-Treat (ITT) and Complier Average Treatment Effects (CACE) of Intervention on Rumor Acceptance in Study 1

	DV: Additive Scale				DV: Gain in Repeated Rumor			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (ITT)	-0.12*** (0.02)				-0.24* (0.03)			
Newspaper (CACE)	-0.87*** (0.17)				-1.79*** (0.35)			
Email (CACE)	-0.24*** (0.04)				-0.50*** (0.07)			
Newspaper/Email (CACE)	-0.21*** (0.03)				-0.44*** (0.06)			
Constant	0.28*** (0.01)	0.28*** (0.01)	0.28*** (0.01)	0.28*** (0.01)	0.17*** (0.03)	0.17*** (0.03)	0.17*** (0.03)	0.17*** (0.03)
Model	OLS	2SLS	2SLS	2SLS	OLS	2SLS	2SLS	2SLS
n	731	731	731	731	731	731	731	731

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.10$ . Robust standard errors are in parentheses. Variables instrumented in 2SLS models: Newspaper, Email, Newspaper/Email.

affect post-treatment levels of political knowledge, attitudes towards fact-checking corrections, or trust in the media.<sup>21</sup>

### Study 2

We conducted a second study using a two-wave online survey combined with a similar experimental intervention during the early months of 2022 in São Paulo. The first wave included 1,037 respondents and was conducted between February 2 and 14. The second wave occurred between February 23 and March 11, and re-interviewed 694 respondents.<sup>22</sup>

Study 2 has a few important differences relative to Study 1. First, the study was not conducted during an electoral campaign. Second, because we did not partner with *Folha de São Paulo* in Study 2, the treatment group received only the email used as part of the treatment in the first study, but not the voucher to subscribe to the newspaper. Third, in order to better parse out the roles of ability and motivation to reject rumors, we also included real news pieces to use as dependent variable. The study included a larger battery of seven false rumors, three of which repeated between waves, and a battery of three factually accurate news pieces, with one repeating between waves. Finally, the programming and randomization of

<sup>21</sup>See Section 10 and Section 11 of supplemental appendix for analyses with alternative dependent variables.

<sup>22</sup>See Section 12 and Section 13 for instrumentation in English and Portuguese.

Study 2 were conducted on a Qualtrics platform, rather than on the polling company's platform, to avoid the imbalances we found in Study 1.<sup>23</sup>

We do not find statistical associations between treatment assignment and pre-treatment covariates in Study 2. A total of 521 subjects from wave 1 were assigned to the treatment group (email), while 516 were assigned to the control group (no message). We also do not find evidence of differential attrition, as the subjects from the experimental groups were equally likely to drop between waves 1 and 2 of the survey.<sup>24</sup> With respect to compliance, a total of 314 subjects from the treatment group opened the email, with 275 of those responding on wave 2. Overall, we were able to reinterview 694 respondents in the second wave (67%).

Table 2 shows ITT and CACE estimates for Study 2 unadjusted for covariates.<sup>25</sup> Similar to study 1, the first dependent variable is the additive scale of rumor acceptance for the seven rumors presented in the second wave of Study 2 (each coded as 0 for rejection and 1 for acceptance). This variable is re-scaled from 0 to 1 and indicates the extent to which subjects accept false rumors as true. The second is the additive scale of acceptance of the three true news pieces we included in Study 2. The third is the difference between the additive scales of acceptance of true and false stories. Columns 1–6 show the estimates for the additive scales described above. Columns 7–12 show the estimates for change in each of the three scales between waves 1 and 2 using only the stories that repeated between waves. This variable takes the value of  $-1$  (believed in wave 1 but not in wave 2),  $0$  (did not change response between waves), and  $1$  (did not believe in wave 1 and believed in wave 2).

For the additive scales in wave 2, the intent-to-treat estimate is statistically significant only at the 10% level ( $p < 0.06$ , column 1). The same is observed for the CACE estimates ( $p < 0.06$ , column 2). The treatment does not affect subjects' acceptance of real news (columns 3 and 4). More notably, the treatment has a significant effect (both ITT and CACE) on the differences in acceptance between real and false stories (columns 5 and 6), which suggests that the intervention does not only reduce acceptance by fostering generalized skepticism towards news content. With respect to the repeated rumors, both ITT and CACE estimates show that the treatment reduces rumor acceptance (columns 7 and 8) while having no effect on true news acceptance (columns 9 and 10), which results in a positive net effect (columns 11 and 12).<sup>26</sup>

All in all, Studies 1 and 2 have similar findings: preemptive interventions are effective at reducing belief in false stories. Although the estimates of Study 2 are smaller and statistically different from the estimates found in Study 1,<sup>27</sup> they are qualitatively similar.<sup>28</sup> Furthermore, Study 2 finding that the intervention did not

<sup>23</sup>See Section 14 of the supplemental appendix for descriptive statistics.

<sup>24</sup>See Section 15 of the supplemental appendix for tests of sample imbalances and section 16 for panel attrition.

<sup>25</sup>See Section 17 for models with covariate adjustment.

<sup>26</sup>See Section 18 of supplemental appendix for bounding estimates.

<sup>27</sup>Cohen's  $d$  for the additive scale is  $-0.15$  [ $-0.30; 0.01$ ] and for the repeated rumors is  $-0.19$  [ $-0.34; -0.03$ ].

<sup>28</sup>Using the pooled average treatment effect (Gerber and Green 2012, 361) to integrate the effect size estimates for the additive scale from studies 1 ( $-0.48$ ) and 2 ( $-0.15$ ) yields a pooled effect size of  $-0.31$  ( $p < 0.01$ ). The combined effect sizes from the two studies ( $-0.31$ ,  $p < 0.01$ ) is larger than the magnitude of effects observed in similar experimental interventions conducted in different contexts (Guess *et al.* 2020).

**Table 2.** Intent-to-Treat (ITT) and Complier Average Treatment Effects (CACE) of Intervention on Rumor Acceptance in Study 2

	Additive Scale						Gain in Repeated Exposure					
	False		True		True-False		False		True		True-False	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment (ITT)	-0.03*		0.04		0.07**		-0.05**		0.04		0.09**	
	(0.02)		(0.02)		(0.03)		(0.02)		(0.04)		(0.04)	
Email (CACE)		-0.04*		0.04		0.08**		-0.06**		0.05		0.11**
		(0.02)		(0.03)		(0.04)		(0.03)		(0.05)		(0.05)
Constant	0.28***	0.28***	0.54*	0.54***	0.25*	0.25***	-0.01	-0.01	-0.05*	-0.05*	-0.04	-0.04
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
Model	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
n	674	674	674	674	674	674	674	674	674	674	674	674

\*\*\* $p < 0.01$ . \*\* $p < 0.05$ . \* $p < 0.10$ . Robust standard errors are in parentheses.

Variable instrumented in 2SLS models: Email.

lead to more skepticism regarding true news stories is consistent with Hypothesis 2 and provides additional insights into the effectiveness of the intervention.

Like in Study 1, we do not find that the treatment affects trust in media.<sup>29</sup> Similar to Study 1, the intervention had a statistically significant effect on rumors in favor of President Bolsonaro, but not on those against him. However, these estimates are not statistically different from each other in Study 2. Also, we do not observe moderation through variables such as education, political interest, media trust, and support for the president, which follows the patterns observed in Study 1.<sup>30</sup>

## Conclusion

Interventions to lower citizens' propensity to accept false rumors can be effective, as our findings suggest. The intervention during the 2020 mayoral elections in São Paulo reduced rumor acceptance among those assigned to the treatment group, and also among those who complied with the attempt to treat. Moreover, the results from the second study show that the intervention reduced belief in fake news without having the same effect for real news.

It is important to note that our design does not directly test the different possible mechanisms by which preemptive interventions inoculate individuals against misinformation. The intervention may reduce misinformation because individuals receive more access to resources that are necessary for identifying false stories, become more motivated to question and investigate their truthfulness, or more skilled in order to detect their false claims. Each dimension has distinct implications for the ways preemptive interventions can be designed. Therefore, future scholarship can contribute to the debate by assessing the effectiveness of distinct types of interventions with respect to how they foster the various possible mechanisms in the political learning process.

**Supplementary material.** To view supplementary material for this article, please visit <https://doi.org/10.1017/XPS.2023.11>

**Data availability statement.** Support for Study 2 was provided by Emory University. The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at <https://doi.org/10.7910/DVN/ZHWIWG>.

**Acknowledgements.** We would like to thank Eduardo Scolese and Fábio Takahashi from Folha de São Paulo and Quaest' operations team for the partnership in Study 1. We are also grateful to Cindy Kam and participants of the INSPER seminar series for their helpful comments and advice and to João Pedro Oliveira dos Santos for the excellent research assistance. Natália S. Bueno is a research associate at FGV/CEPESP and she is grateful for FGV's support in conducting this study.

**Competing interest.** One of the authors (Felipe Nunes) is also the CEO of *Quaest Consultoria & Pesquisa*, the company that conducted the two studies presented in the paper. Study 1 resulted from a partnership between *Quaest* and the newspaper *Folha de São Paulo*. It was fully funded by *Quaest*. *Quaest* did not receive

<sup>29</sup>We also included questions for self-reported attention and ability to identify fake news as an attempt to parse out the mechanisms, but the results are inconsistent and affected by differential panel attrition based on the two variables. See Section 19 of the supplemental appendix for results.

<sup>30</sup>See Section 20 of supplemental appendix for estimates of heterogeneous effects and Section 21 alternative dependent variables.

any payment from Folha de São Paulo or the universities involved for conducting the two waves of the study. Felipe Nunes did not receive any financial return from this research either. The partnership between *Quaest* and the newspaper *Folha de São Paulo* did not happen in Study 2. Study 2 was fully funded by Research Funds from Emory University.

**Ethics statement.** This research adheres to APSA's Principles and Guidance for Human Subjects Research. Both studies were reviewed and approved by Institutional Review Boards in two U.S. institutions (UNCC and Emory University) and one Brazilian institution (Fundação Getúlio Vargas). In any case, exposing subjects to misinformation raises ethical issues. We addressed these concerns by taking the following precautions: we used false stories that had large circulation in Brazil and were not fabricated by the researchers and all participants were debriefed at the end of the study. For more information on subjects and recruitment process, see Section 22 of the Supplemental Appendix. A consort flow diagram is included in Section 23 of the Supplemental Appendix.

Importantly, all subjects consented to take the survey, and it did not contain deception. The researchers did not have access to identifiable data.

## References

- Angrist, J. D., G. W. Imbens, and D. B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association* 91(434): 444–55.
- Badrinathan, S. 2021. Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. *American Political Science Review*, 115(4): 1325–1341. doi: [10.1017/S0003055421000459](https://doi.org/10.1017/S0003055421000459)
- Banas, J. A. and G. Miller. 2013. Inducing Resistance to Conspiracy Theory Propaganda: Testing Inoculation and Metainoculation Strategies. *Human Communication Research* 39: 184–207.
- Batista Pereira, F., N. S. Bueno, F. Nunes, and N. Pavão. 2022. Fake News, Fact Checking, and Partisanship: The Resilience of Rumors in the 2018 Brazilian Elections. *The Journal of Politics* 84(4): 2188–201.
- Batista Pereira, F., N. S. Bueno, N. Pavão, and F. Nunes. 2023. Replication Data for: Inoculation Reduces Misinformation: Experimental Evidence from Multidimensional Interventions in Brazil. doi: [10.7910/DVN/ZHWIWG](https://doi.org/10.7910/DVN/ZHWIWG).
- Bulger, M. and P. Davidson. 2018. The Promises, Challenges and Futures of Media Literacy. *Journal of Media Literacy Education* 10(1): 1–21.
- Carey, J. M., V. Chi, D. J. Flynn, B. Nyhan, and T. Zeitoff. 2020. The Effects of Corrective Information about Disease Epidemics and Outbreaks: Evidence from Zika and Yellow Fever in Brazil. *Science Advances* 6(5): eaaw7449. doi: [10.1126/sciadv.aaw7449](https://doi.org/10.1126/sciadv.aaw7449)
- Clayton, K., S. Blair, J. A. Busam, S. Forstner, J. Glance, G. G. A. Kawata, A. Kovvuri, J. Martin, E. Morgan, M. Sandhu, R. Sang, R. ScholzBright, A. T. Welch, A. G. Wolff, A. Zhou, and B. Nyhan. 2020. Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Factcheck Tags in Reducing Belief in False Stories on Social Media. *Political Behavior* 42: 1073–95.
- Cook, J., S. Lewandowsky, and U. K. H. Ecker. 2017. Neutralizing Misinformation Through Inoculation: Exposing Misleading Argumentation Techniques Reduces Their Influence. *PLOS One* 12(5): e0175799.
- Eagly, A. H. and S. Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.
- Fazio, L. K., D. G. Rand, and G. Pennycook. 2019. Repetition Increases Perceived Truth Equally for Plausible and Implausible Statements. *Psychonomic Bulletin & Review* 26: 1705–10.
- Gerber, A. and D. P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York, NY: W. W. Norton & Company.
- Guess, A. M., M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar. 2020. A Digital Media Literacy Intervention Increases Discernment between Mainstream and False News in the United States and India. *Proceedings of the National Academy of Sciences* 117(27): 15536–45.
- Hameleers, M. 2022. Separating Truth from Lies: Comparing the Effects of News Media Literacy Interventions and Fact-Checkers in Response to Political Misinformation in the US and Netherlands. *Information, Communication & Society* 25(1): 110–126. doi: [10.1080/1369118X.2020.1764603](https://doi.org/10.1080/1369118X.2020.1764603)

- Jeong, S.-H., H. Cho, and Y. Hwang.** 2012. Media Literacy Interventions: A Meta-Analytic Review. *Journal of Communication* 62: 454–72.
- Jones-Jang, S. M., T. Mortensen, and J. Liu.** 2021. Does Media Literacy Help Identification of Fake News? Information Literacy Helps, But Other Literacies Don't. *American Behavioral Scientist* 65(2): 371–88.
- Lee, N. M.** 2018. Fake News, Phishing, and Fraud: A Call for Research on Digital Media Literacy Education Beyond the Classroom. *Communication Education* 67(4): 460–66.
- Lutzke, L., C. Drummond, P. Slovic, and J. Árvai.** 2019. Priming Critical Thinking: Simple Interventions Limit the influence of fake news about climate change on facebook. *Global Environmental Change* 58: 101964.
- Machado, C., B. Kira, V. Narayanan, B. Kollanyi, and P. N. Howard.** 2019. A Study of Misinformation in Whatsapp Groups with a Focus on the Brazilian Presidential Elections. *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, 1013–1019. doi: [10.1145/3308560.3316738](https://doi.org/10.1145/3308560.3316738)
- McGuire, W. J.** 1961. The Effectiveness of Supportive and Refutational Defenses in Immunizing and Restoring Beliefs against Persuasion. *Sociometry* 24(2): 184–97.
- McIntery, L.** 2018. *Post-Truth*. Cambridge, MA: MIT Press.
- Pennycook, G., T. D. Cannon, and D. G. Rand.** 2018. Prior Exposure Increases Perceived Accuracy of Fake News. *Journal of Experimental Psychology: General* 147(12): 1865–80.
- Pennycook, G. and D. G. Rand.** 2019. Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning. *Cognition* 188: 38–50.
- Petty, R. E. and J. T. Cacioppo.** 1996. *Attitudes and Persuasion: Classic And Contemporary Approaches*. New York, NY: Avalon Publishing.
- Porter, E., Y. Velez, and T. J. Wood.** 2023. Correcting COVID-19 vaccine misinformation in 10 countries. *Royal Society Open Science* 10(3): 221097. doi: [10.1098/rsos.221097](https://doi.org/10.1098/rsos.221097)
- Rozenbeek, J. and S. van der Linden.** 2019. Fake News Game Confers Psychological Resistance against Online Misinformation. *Palgrave Communications* 5(65): 1–10.
- Rossini, P., J. Stromer-Galley, E. A. Baptista, and V. V. de Oliveira.** 2021. Dysfunctional Information Sharing on Whatsapp and Facebook: The Role of Political Talk, Cross-Cutting Exposure and Social Corrections. *New Media & Society* 23(8): 2430–2451. doi: [10.1177/1461444820928059](https://doi.org/10.1177/1461444820928059)
- Tully, M., E. K. Vraga, and L. Bode.** 2020. Designing and Testing News Literacy Messages for Social Media. *Mass Communication and Society* 23(1): 22–46.

---

**Cite this article:** Batista Pereira F, Bueno NS, Nunes F, and Pávao N. Inoculation Reduces Misinformation: Experimental Evidence from Multidimensional Interventions in Brazil. *Journal of Experimental Political Science*. <https://doi.org/10.1017/XPS.2023.11>