"I Don't Think That's True, Bro!"

Social Corrections for Misinformation in India

Sumitra Badrinathan
University of Oxford

Simon Chauchard
Leiden University *

Abstract

Fact-checks and corrections of falsehoods have emerged as effective ways to counter misinformation online. But in contexts with encrypted messaging applications (EMAs), corrections must necessarily emanate from peers. Are such social corrections effective? If so, how substantiated do corrective messages need to be? To answer these questions, we evaluate the effect of different types of social corrections on the persistence of misinformation in India (N=5100). Using an online experiment, we show that social corrections substantially reduce beliefs in misinformation, including in beliefs deeply anchored in salient group identities. Importantly, these positive effects are not attenuated by partisan motivated reasoning, highlighting a striking difference from Western contexts. We also find that the presence of a correction matters more than how sophisticated this correction is: substantiating a correction with a source only improves its effect in a minority of cases; besides, when social corrections are effective, citing a source does not drastically improve the size of their effect. These results have implications for both users and platforms and speak to countering misinformation in developing countries that rely on private messaging apps.

Keywords: Misinformation; Social Media; Peer Correction; Motivated Reasoning; WhatsApp; India

*This study was registered with Evidence in Governance and Policy (20191008AB) and received IRB approval from Columbia University and IE University (IRB-AAAS3860). The authors thank Ipsa Arora, Ritubhan Gautam, and Hanmant Wanole for research assistance. This research was funded by Facebook. The authors thank Alex Leavitt and Devra Moelher, as well as participants at the Facebook Integrity Research Workshop (June 2019). For comments on the manuscript, the authors are grateful to seminar participants at Columbia University, Leiden University, and the American Political Science Association conference.

Introduction

Over the past decade, a vast research agenda has tested the effect of corrective messages to fight misinformation on social media. The majority of such studies suggest that corrective interventions lead to small and beneficial effects (Chan et al. 2017; Fridkin, Kenney, and Wintersieck 2015; Porter and Wood 2021). Since users sometimes also correct each other, a part of this literature has, in addition, explored the effect of "social corrections" and found them to be similarly efficient (Bode and Vraga 2018; van der Meer and Jin 2020; Martel, Mosleh, and Rand 2021).

However, these encouraging findings about the effect of corrections on changing misinformed beliefs remain largely limited to Western democracies and Global North contexts, with little scholarship, by contrast, focusing on misinformation in the developing world (Rossini et al. 2020; Pereira et al. 2021; Rosenzweig et al. 2021; Guess et al. 2020; Neyazi, Kalogeropoulos, and Nielsen 2021). Countries in the Global South are thus not only understudied, but misinformation in these contexts can have disastrous offline consequences such as fueling political violence and ethnic riots.

In this article, we build on the emerging literature on social corrections – i.e. peer-to-peer corrective messages – and explore the extent to which such corrections are effective in contexts in which they are most sorely needed. In countries like India, pernicious misinformation builds on longstanding ethnic fractures, often times resulting in conflict and violence. Further, since a large part of this misinformation circulates through encrypted messaging applications (EMAs) on which corrections can *only* come from other users (as opposed to platform-based warnings and labels), it is especially crucial to understand whether social corrections can play a role in combating misinformation in such understudied contexts.

Are social corrections effective in such contexts? If so, what type of corrective messages work best at altering misinformed beliefs? To answer these questions, this study tests the effect of the presence of a social correction and the type of correction on belief

in electoral misinformation in India. We implement a large-scale online experiment with over 5,100 respondents in the aftermath of the contentious 2019 general elections, which saw a dramatic surge in political misinformation on social media. We show respondents a series of hypothetical WhatsApp group chat conversations. In each conversation, a user posts a false story to which a peer reacts, with or without a correction, and with or without citing evidence.

Results demonstrate that the presence of a social correction significantly reduces belief in misinformation. Relative to a no correction condition, witnessing *any* type of social correction reduces the perceived accuracy of beliefs in misinformation. Importantly, this effect is robust to respondents' partisan identity and persists across different types of misinformation stories, including deep-rooted beliefs. We find that partisan motivated reasoning does not attenuate these corrective effects: our treatments are just as likely to work on partisans from both sides of the ideological spectrum, suggesting important differences in the mechanisms through which misinformation spreads in contexts like India. Finally, we show that the source and the sophistication of corrective messages do not strongly condition their effect: in our experiment, brief, unsourced and unsubstantiated corrective messages – which may be thought as more realistic corrections on EMAs – often perform just as well relative to corrections citing evidence from a variety of credible sources.

We note two key implications of our findings. On the one hand, our results suggest that social corrections have a positive effect in the Indian context. Hence, incentivizing users to engage in speaking up, correcting misinformation, or verbalizing their fact-checks on homogeneous networks may help in the reduction of misinformed beliefs. On the other hand, because respondents emerge – in this context – as influenceable (minimal corrective messages are relatively effective), encouraging users to sound off as easily as possible may also have perverse consequences if bad faith actors themselves use such strategies to pedal more misinformation. This study hopes to spark a robust

research agenda on solutions to misinformation in the Global South, focusing on the challenges that encryption, low digital literacy, and social cleavages bring to information processing.

Social Corrections Across Platforms and Contexts

Social corrections have become a key element in combating misinformation, with platforms increasingly encouraging users to check on each other. This is especially relevant to EMAs: in 2019, faced with public outcry over political misinformation going viral on the platform in the lead-up to a general election, WhatsApp launched a large-scale ad campaign to encourage its Indian users to check on their peers' claims, and if possible, to use fact-checks produced by affiliated organizations in order to do so.[1] If users indeed followed through with this recommendation and fact-checked their peers, to what extent should we expect such a strategy to reduce the uptake of misinformation in contexts like India?

Existing research focused on the American context suggests that such social corrections might indeed be beneficial. Bode and Vraga (2018) compare algorithmic corrections with peer corrections and find that they are both equally effective at dispelling misinformation. Building on the same idea, van der Meer and Jin (2020) demonstrate that peer corrections are effective at reducing misinformation relative to a control condition.

The literature on source credibility provides a mechanism explaining why social corrections may be effective. Individuals have limited time and cognitive resources to comprehend complex topics such as policy or current affairs, and may therefore use the perceived credibility of sources as a heuristic to guide their evaluation of what is true or false. In general, high-credibility sources are more persuasive and promote greater atti-

[1] See, for instance, the advertising widely distributed across the Indian press in Appendix A.

tude change than low credibility sources (Eagly and Chaiken 1993). Further, while both expertise and trustworthiness are components of source credibility, the latter is found to be more effective in persuasion than the former (Swire and Ecker 2018). Thus, arguably, peers should be seen as more trustworthy than unknown or distant individuals, and users on social media are likely to be able to persuade their peers. Additionally, homophily, or the extent to which a person perceives similarities between the way they think and another person does, is often seen as a key determinant of source credibility (Housholder and LaMarre 2014). Networks on social media, including in India, are likely to be comprised of likeminded individuals who share political and other views (Tokita, Guess, and Tarnita 2021). Following this argument, it is likely that peers would constitute credible (and hence persuasive) sources in this context. Accordingly, we hypothesize that:

Hypothesis 1: Exposure to corrective messages emanating from peers will reduce the perceived accuracy of misinformation, relative to a no correction condition.

Despite the relative success of fact-checking efforts, the empirical literature on misinformation demonstrates that the success of corrections is a function of individuals' preexisting beliefs. In this regard, a primary factor influencing the efficacy of corrections is partisan motivated reasoning (Thorson 2016; Flynn, Nyhan, and Reifler 2017).

Existing findings on partisan motivated reasoning mainly come from the American context, where partisan polarization notoriously acts as a perceptual screen (Green, Palmquist, and Schickler 2004). In India, however, that partisan motivated reasoning will affect corrections is not a foregone conclusion (Badrinathan 2021). On the one hand, partisan affiliations in India have been shown to be traditionally weaker and less stable, sometimes forming for non-ideological reasons (Chandra 2004; Bussell 2019). On the other hand, reports show that misinformation is largely political on WhatsApp, with groups formed by the ruling Bharatiya Janta Party (BJP), the right-wing, Hindu nationalist government of India, often morphing into havens of misinformation and hateful

rhetoric, capable of inciting violence (Arun 2019; Farooq 2017). Indeed, research on WhatsApp groups demonstrates that a majority of political content comes from groups allied with the BJP (Garimella and Eckles 2020). Further, since we conduct this experiment after a contentious election, during which attachments to parties were arguably heightened (Michelitch and Utych 2018), we might expect motivated reasoning to play a role in information processing.

Despite these contrasting priors, in keeping with findings from the literature on fact-checking we hypothesize that partisan motivated reasoning should attenuate the effects of corrective messages (Taber and Lodge 2006). Specifically, we hypothesize that both the political slant of misinformation, as well as the news source reporting it, can condition the effectiveness of corrections:

Hypothesis 2: Peer corrections will be more effective when misinformation is attributed to an ideologically dissonant politician (compared to when it is unattributed).

Hypothesis 3: Peer corrections will be more effective when misinformation originates from a dissonant media outlet (compared to unattributed or neutral outlet).

Hypothesis 4: Peer corrections will be less effective when the slant of the story is ideologically congruent (compared to non-ideological stories).

Finally, we also consider whether corrections backed by evidence are more or less effective relative to one-line corrections. Research demonstrates that the persuasive quality of an argument is a function of whether or not it is substantiated (Stiff and Mongeau 2016; German et al. 2016; Reinard 1988). With social media, Vraga and Bode (2018) test the effect of social corrections on Facebook and Twitter and find that corrections substantiated with a source are more effective at countering misinformation.

Thus, the level of substantiation of corrective messages should matter for reducing the uptake of misinformation. In this study we distinguish between substantiated corrective messages, or those that are backed by an explanation or source, and unsubstantiated corrections which have no source or explanation, and are consequently shorter. We thus

hypothesize:

Hypothesis 5: Exposure to unsubstantiated corrections (relative to substantiated corrections) will be less effective at countering misinformation.

Design

To test these hypotheses, we designed and fielded an online experiment in India (N $\approx$ 5,100) in the aftermath of the contentious 2019 general elections. In our experiment, respondents were recruited through Facebook and randomly assigned to one of four conditions. In all conditions, respondents were shown (in random order) a series of nine hypothetical conversations on WhatsApp group chats, seven of which contained misinformation.

The theme of these nine conversations – and hence of the misinformation stimulus respondents were shown – was chosen following a pretest with an online Indian sample. Each of the misinformation stories selected for the final experiment was strongly believed or believed by at least of 25% of the pretest sample, with some of these statements being believed by a large majority of respondents.[2]

The final selection of stories was the product of several constraints and choices. To avoid prompting respondents to systematically reject the veracity of rumors, we included some true stories (2 out of 9). But simultaneously, our goal was to maximize respondent exposure to corrections for controversial fake political rumors that spread widely during the run up to the 2019 elections in India, hence our distribution skewed in favor of false stories (7 out of 9). We selected stories encompassing a broad variety of topics including current electoral politics (stories 8 and 9), health (stories 5 and 6), historical conspiracies (story 7), and in order to test the effect of social corrections on identity-related misinformation, religion and minorities (stories 3 and 4).

[2]Data from the pretest is presented in Appendix I.

These stories are listed in Table 1.

Table 1: Dependent Variable Stories

	Story	Veracity
1	Australia is the country that has won the ICC cricket world cup the most often	True
2	There is no cure for HIV/AIDS	True
3	In the future, the Muslim population in India will overtake the Hindu population in India	False
4	Polygamy is very common in the Muslim population	False
5	M-R vaccines are associated with autism and retardation	False
6	Drinking cow urine (gomutra) can help build one's immune system	False
7	Netaji Bose did not die in a plane crash in 1945	False
8	The BJP has hacked electronic voting machines	False
9	UNESCO declared PM Modi best Prime Minister in 2016	False

Experimental Groups

Respondents were randomly assigned to one of four experimental groups including three treatment and one control group. Control group respondents read a WhatsApp conversation that included a misinformation stimulus posted by a hypothetical user. In the three treatment conditions, the misinformation stimulus was followed by a peer correction posted by a second user.

Since our goal was to test the efficacy of different types of corrections, our three treatment groups varied the degree of substantiation of the corrective message, as well as its source. Respondents in the *Domain Expert* treatment read a substantiated correction pointing to a domain expert as the source of the correction (for example, the Election Commission of India for electoral misinformation, or the Census Bureau of India for demographic misinformation). Respondents in the *Fact Checker* treatment read a substantiated correction pointing to a verified fact-checker in India as having debunked the misinformation posted.[3] Respondents in the *Unsubstantiated Correction* treatment, on

[3]We further randomized the specific fact-checker such that respondents read a fact-check from one of five sources: the online platform WhatsApp itself, the online platform Facebook, India's oldest print newspaper The Times of India, left-leaning third party fact-checking service AltNews, or right-leaning

the other hand, read a correction that was a simple rebuttal by the second user, devoid of substantiation or a source of correction. This included a one-line simple correction (for instance saying "I don't think that's true, bro!") in response to the misinformation stimulus. The full text of each experimental manipulation, along with samples of the experimental stimuli, is included in Online Appendix C.

In addition to experimentally varying the presence and the degree of sophistication of the correction, we test for partisan motivated reasoning by varying the news outlet / source reporting the false story,[4] and/or the political identity of the politician from whom the misinformation originated.[5]

For each story, respondents were equally likely to be randomized into one of the four experimental groups. Further, within each treatment condition respondents had an equal probability of being assigned to each of the possible combinations of media outlets x politicians listed above.[6]

Procedure and Outcome Variable

After reading each WhatsApp conversation, respondents in all four groups were asked to evaluate the veracity of the misinformation stimulus included the conversation with a single outcome question:

How accurate is the following statement? [Statement of the rumor]

(very accurate, somewhat accurate, not very accurate, not at all accurate)

third party fact-checking service Vishwas News. We explore the differences between these sources cited in a related working paper.

[4]We vary the media outlet reporting the story to include three possible sources: India TV (a relatively pro-BJP private news channel), NDTV Hindi (a relatively anti-BJP private channel), and DD News (public channel).

[5]To avoid biasing responses, we deliberately blacked out the purported names of the participants and presented this as a measure to protect their privacy, hence likely increasing the realism of the experiment. We explore the effect of these manipulations in related work.

[6]A small proportion of respondents (3% of the sample) were randomized into a "pure control" condition in which we measure the dependent variable of belief in misinformation without showing respondents any of the screenshots. As demonstrated in Appendix Table J.1 we detect no statistical differences in the overall rate of belief in all false stories when comparing our pure control condition to control condition with no correction, though we cannot exclude the possibility that these would be underpowered analyses.

Our design took several steps to increase external validity and realism. First, as noted above, we selected a diverse sample of stories. Further, we excluded highly unrealistic manipulations and tailored domain expert corrections to each rumor (e.g., we attribute expert corrections of voter fraud rumors to the Election Commission of India). Finally, given that respondents each saw nine screenshots, we slightly varied the specific text of the messages in each screenshot to ensure realism. Online Appendix C gives a full list of treatments for each rumor used in our experiment and provides a sample of WhatsApp screenshots shown to treatment group respondents.

Sample

Participants in this study were Hindi speakers recruited through Facebook. The ad used to recruit respondents is presented in Online Appendix B. To be eligible to participate, respondents were required to be adult residents of India who used WhatsApp. While we recruited over 5100 participants, the actual N presented in our analyses varies slightly for each dependent variable story (+ or - 1%), as we include observations from respondents who exit the survey before reading all 9 screenshots.[7] The experiment was conducted entirely in Hindi. Sample characteristics of our respondents are available in Appendix M.

Results

Figure 1 lists the 7 false stories used as part of the dependent variable measure in this study. This figure plots the share of respondents in the sample who believed each story to be true. Our findings demonstrate the high salience of false stories in the Indian context. Despite the fact that 75% of all screenshots (across conditions) contained a correction, 6 of the 7 false rumors were rated as accurate or somewhat accurate by over

[7] Only 75 respondents, or less than 1.5% of the sample dropped out through the course of the experiment.

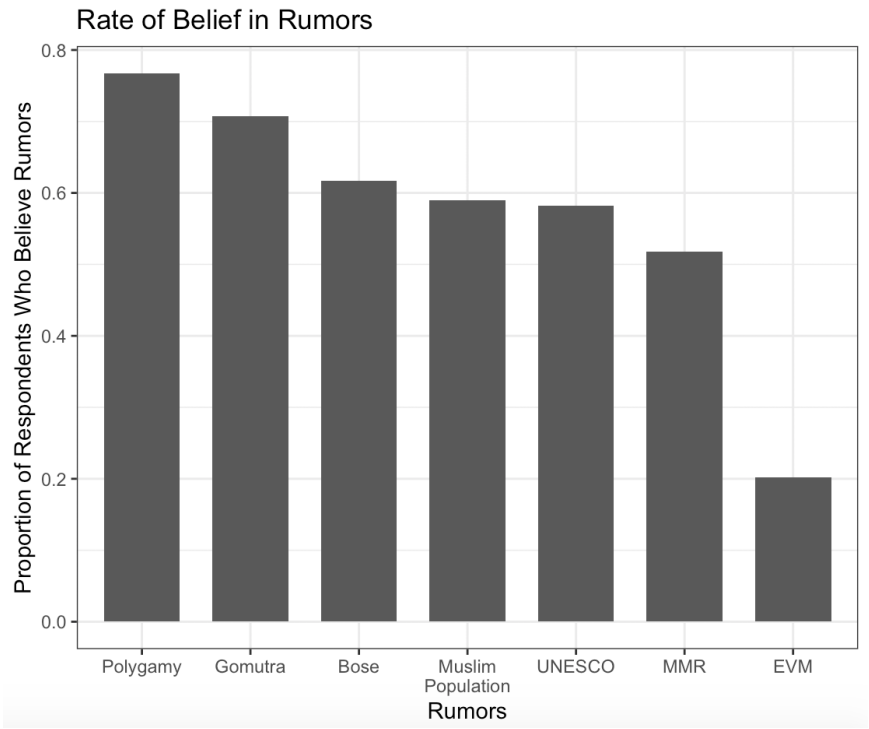


Figure 1: Overall rate of belief in misinformation across experimental conditions

50% of the sample, with the top two prevalent rumors believed by over 75% of the sample, underscoring the tenacity of misinformation in the Indian context.

The Effect of Social Corrections

Do corrections impact these high rates of belief? We first present results for Hypothesis 1, which tests whether exposure to any correction reduces the perceived accuracy of misinformation. To test Hypothesis 1, we pool together all the different types of social corrections such that the primary comparison of interest is between having received a correction (of any kind) and not having received one. This comparison is expressed in Equation 0.1:

$$Belief\,Accuracy_i = \alpha + \beta_1 AnyCorrection_i + \epsilon_i \tag{0.1}$$

In the equation, the *AnyCorrection* variable represents pooled assignment all social

correction conditions (both substantiated and not, relative to control). The dependent variable *BeliefAccuracy* measures the self-reported accuracy rating that respondents give to each story on a 4-point scale, with higher values representing greater perceived accuracy. For more transparency, since our design amounts to running 7 experiments in randomized order, we estimate here a separate bivariate OLS model for each of the seven false stories in our experiment, represented by the seven columns in Table 2.[8]

Our results demonstrate that corrections are effective at reducing beliefs in false stories. Exposure to social corrections significantly reduces the likelihood that respondents report false rumors to be accurate, relative to not receiving any correction. We do not obtain a significant result for only one (out of 7) false story, the rumor that electronic voting machines (EVMs) were hacked by the BJP ahead of the elections. As Figure 1 demonstrates, belief in this rumor was low to begin with, possibly making it harder for the treatment to have an impact. In contrast, a consistent negative effect appears for the remaining stories, although effect sizes vary across rumors. Particularly, we see effects of larger magnitude (greater than 0.4 on a scale from 1 to 4) on two of the stories: the MMR vaccine rumor and the UNESCO rumor.[9] Thus our results show that social corrections, the only suitable techniques for private online spaces, significantly reduce overall rates of beliefs in patently false rumors circulating on WhatsApp in India.

[8]Note that we omit the pure control from these analyses. Combining pure control with the control condition does not change our results, as we demonstrate in Appendix Table J.1 that there is no difference between the control and pure control conditions.

[9]The size of the effect across rumors does not appear related to the prior salience of these rumors in our sampled population. As shown in Appendix Figures I.1 and I.2, many respondents in our pretest had heard of the widely circulated UNESCO rumor, while fewer had heard of the rumor about MMR vaccines. Yet both led to comparatively large corrective effects.

Table 2: Main Effect of Any Correction

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Correction	−0.104***	−0.190***	−0.448***	−0.106***	−0.024	−0.411***	−0.109***
	(0.037)	(0.030)	(0.033)	(0.033)	(0.032)	(0.040)	(0.031)
Constant	2.746***	3.272***	2.827***	3.010***	1.650***	2.999***	2.788***
	(0.033)	(0.026)	(0.028)	(0.027)	(0.027)	(0.034)	(0.025)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.002	0.008	0.035	0.002	0.0001	0.021	0.002
Adjusted R^2	0.001	0.008	0.035	0.002	−0.0001	0.020	0.002
Res. Std. Er.	1.095	0.946	1.039	1.060	1.014	1.251	1.048
F Statistic	7.859***	39.895***	185.869***	10.536***	0.568	107.789***	12.390***

Note: *p<0.1; **p<0.05; ***p<0.01

This result holds in the presence of added control variables. In Appendix Table K.1, we control for the media source reporting the story as well as the politician it is attributed to, two factors likely to impact beliefs. We find that our results are robust to these controls: regardless of the the media source or politician being congruent or dissonant, results hold. Our experimental corrections significantly improve the perceived accuracy of beliefs, and the magnitude of these effects remains relatively unchanged.[10]

Motivated Reasoning and Social Corrections

To what extent are the corrective effects obtained above affected by motivated reasoning in the Indian context? To answer this question, we look at motivated reasoning in three ways. First, we examine whether the partisan identity of the politician making the false claim affects belief in the story. In India, as is often the case in several other contexts, misinformation can emanate from elites who amplify false stories for political gain. Misinformation coming from the top can often have a stronger effect on respondents who are

[10]Given that we report highly significant results ($p < 0.001$ on 3 of the 6 significant coefficients, and $p < 0.005$ on three more) in Table 2, we test whether these results are robust to a Benjamini-Hochberg adjustment. We show that our results remain significant when we perform a Benjamini-Hochberg adjustment fixing the false discovery rate at 5% (see Appendix N).

ideologically inclined with elites making false claims, relative to misinformation coming from non-elites (Van Duyn and Collier 2019). Hence it stands to reason that congruence between respondents' partisan identity and the identity of the politician to whom a story is attributed can impact beliefs.

To test Hypothesis 2, we limit our analyses to the subset of stories that are clearly partisan in nature (Rumors 3, 4, 6, 8, and 9) and code whether the rumor was attributed in the experimental prompt to a copartisan or outpartisan politician. We code a politician as copartisan or outpartisan as a function of the respondent's self-reported partisan inclination towards the ruling party, the BJP, relying on the respondent's expressed closeness to this party. Concretely, a BJP politician is deemed copartisan if the respondent describes themselves as close or very close to the party, and outpartisan if the respondent describes themselves as far or very far from it.

Second, we look at the effect that congruent or dissonant media outlets reporting a story can have on beliefs (Hypothesis 3). To test Hypothesis 3, we code a media outlet as congruent or dissonant as a function of the respondent's expressed proximity to the BJP. Concretely, we code the pro-BJP outlet (here, India TV) as congruent and the anti-BJP outlet (here, NDTV) as dissonant when the respondent reports feeling close or very close to the BJP. By contrast, we code the pro-BJP outlet (India TV) as dissonant and anti-BJP outlet (NDTV) as congruent when the respondent reports feeling far or very far to the BJP.

Finally, we look at the degree of congruence of the story slant itself with respondents' beliefs. To what extent does the efficacy of the treatment depend on whether the content of the false story was congruent with respondents' political beliefs (Hypotheses 4)? To test Hypothesis 4, we again limit our analyses to the subset of rumors that are clearly political (rumors 3, 4, 6, 8,and 9). We code rumors as congruent or dissonant ex-ante as a function of participants own ideological inclinations and as a function of our observations of the two parties' campaign platforms. When participants self-report

being close or very close to the BJP, Rumors 3 (Muslim population growth), 4 (polygamy in the Muslim population), 6 (belief about the virtues of cow urine), and 9 (Modi and UNESCO) are coded as congruent rumors. By contrast, Rumor 8 (EVMs) is coded as dissonant, while Rumors 1, 2, 5 and 7 are coded as neither congenial nor dissonant. Similarly, when participants self-report being close or very close to the Congress Party, the main opposition in India, Rumor 8 is coded as congenial while Rumors 3, 4, 6, 9 are coded as dissonant and Rumors 1, 2, 5, and 7 are neither congenial nor dissonant. [11]

For these hypotheses, our quantity of interest is the interaction between exposure to a corrective message (pooling across all types of corrective messages) and the congeniality or congruence of the source/media outlet/rumor.

Tables 3 and 4 test whether the effect of the correction is a function of the slant of the story itself. While Table 3 looks at whether corrections are less effective for ideologically congruent stories, Table 4 looks are whether corrections are more effective for ideologically dissonant stories (Hypothesis 4).[12]

[11] Results in Appendix Tables G.1 and G.2 underscore this coding choice: we show that rumor congruence significantly predicts higher rates of belief in rumors. Rumors rated in the pretest as congenial are more likely to be believed, while rumors rated as dissonant are less likely to be believed.

[12] In the interest of space, we present results for Hypotheses 2 and 3 in Appendices D and E, respectively.

Table 3: Effect of Correction * Congruent Claim on Belief in Rumor

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
AnyCorrection	−0.092	−0.163***	−0.117**	0.0002	−0.069
	(0.059)	(0.048)	(0.052)	(0.038)	(0.053)
CongruentClaim	0.238***	0.246***	0.369***	0.520***	0.362***
	(0.067)	(0.053)	(0.055)	(0.056)	(0.057)
AnyCorrection*	−0.025	−0.043	0.014	−0.057	0.023
CongruentClaim	(0.076)	(0.061)	(0.066)	(0.066)	(0.067)
Constant	2.602***	3.120***	2.784***	1.478***	2.751***
	(0.052)	(0.041)	(0.043)	(0.032)	(0.045)
Observations	5,104	5,103	5,099	5,136	5,099
R^2	0.011	0.020	0.032	0.049	0.031
Adjusted R^2	0.010	0.019	0.032	0.049	0.030
Res. Std. Er.	1.090 (df = 5100)	0.940 (df = 5099)	1.044 (df = 5095)	0.989 (df = 5132)	1.045 (df = 5095)
F Statistic	18.919***	34.488***	56.388***	88.471***	53.490***

Note: *p<0.1; **p<0.05; ***p<0.01

Across Tables 3 and 4, the interaction between the treatment (any correction) and the slant of the story corrected produces a null result. The results from these tests thus point to a striking conclusion: the effect of corrections is *not* limited by the ideological slant of the story.

Similar results emerge when we look at motivated reasoning in two other ways, as per the politician to whom a story is attributed, or the news outlet reporting the story. We find that interacting the correction with the identity of the politician or media outlet does not reduce or change the effect of corrections. Results from these tests are reported in Appendix Tables D.1 and D.2 (effect of the identity of the politician) and Tables E.1 and E.2 (effect of media outlet).

Table 4: Effect of Correction * Dissonant Claim on Belief in Rumor

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
AnyCorrection	−0.127***	−0.195***	−0.088**	−0.004	−0.032
	(0.045)	(0.036)	(0.039)	(0.049)	(0.040)
DissonantClaim	−0.206***	−0.188***	−0.267***	−0.608***	−0.267***
	(0.069)	(0.055)	(0.058)	(0.053)	(0.060)
AnyCorrection*	0.062	0.016	−0.060	−0.022	−0.058
DissonantClaim	(0.078)	(0.064)	(0.069)	(0.062)	(0.070)
Constant	2.816***	3.334***	3.097***	2.022***	3.058***
	(0.040)	(0.031)	(0.033)	(0.042)	(0.035)
Observations	5,104	5,103	5,099	5,136	5,099
R^2	0.006	0.015	0.021	0.090	0.019
Adjusted R^2	0.006	0.015	0.020	0.089	0.019
Res. Std. Er.	1.093 (df = 5100)	0.942 (df = 5099)	1.050 (df = 5095)	0.968 (df = 5132)	1.051 (df = 5095)
F Statistic	10.658***	26.475***	36.056***	168.534***	33.051***

Note: *p<0.1; **p<0.05; ***p<0.01

Taken together, these results demonstrate that the effectiveness of corrections persists despite partisan ties or partisan motivated reasoning, and that corrective effects in this context are *not* conditional on the source or slant of a rumor. Importantly, our results exclude the possibility that these results might owe to our sample being a low-effort sample: as visible in tables 3 and 4, our respondents did strongly react - in the expected direction - to the slant of rumors, to their cited sources and to the presence of a correction. This, however, did not mitigate their reaction to the correction. Besides, while our sample tilts towards BJP supporters, educated, upper-caste men (and as such reflecting the population of high-frequency social media users in India), the relatively large size of this sample (N>5,000) makes it unlikely that these results owe to insufficient statistical power. These factors suggest that partisan motivated reasoning plays a comparatively less important role in India, breaking with results frequently obtained in the American context (Nyhan and Reifler 2010).

An additional finding confirms this interpretation. We test whether partisans of

the ruling party (the BJP) are susceptible to partisan motivated reasoning. Since there is a supply side bias in political motivation in India (with misinformation often emanating from BJP-sympathetic sources) and since we code party identity as feelings towards the BJP, it stands to reason that BJP party identity is more consolidated in our sample (those not supporting the BJP may support a host of other national and regional parties in the country). We find that even among this subgroup of arguably stronger and more consolidated partisans, no motivated reasoning effect exists (see Appendix Table F.1). This finding persists when we run a second series of tests relying on reported voting decisions in the 2019 elections instead of measuring respondents' closeness to the BJP: participants who voted for the ruling party in the 2019 election do not react to corrections any differently. These findings underscore the relative absence of partisan motivated reasoning in the Indian context. We explore this result further in the discussion.

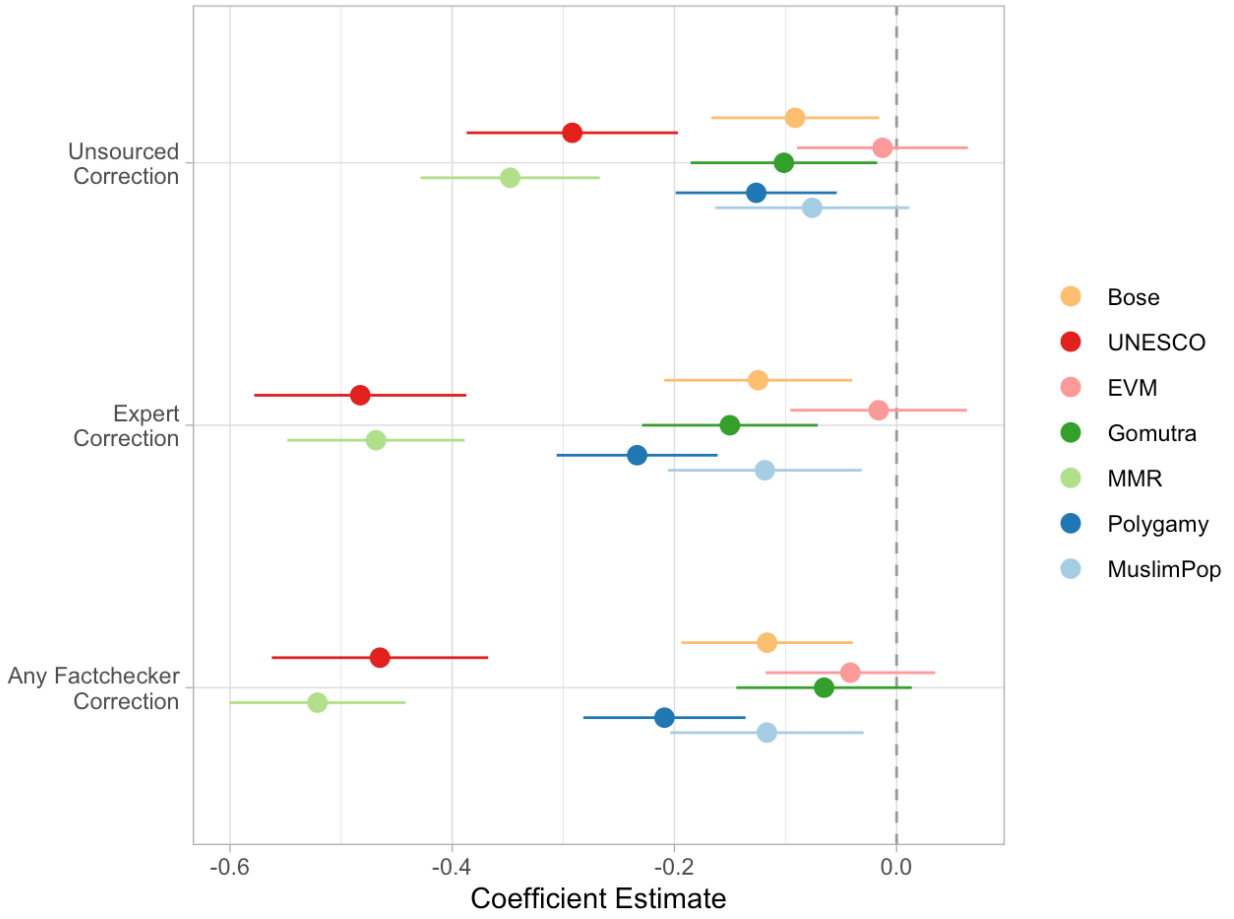
Are Substantiated Corrections More Effective?

Our main effect in this paper demonstrates that receiving *any* correction (relative to control) can reduce the perceived accuracy of misinformation. While this analysis pools together all corrections, we now examine which types of corrections are most effective. Particularly, we compare substantiated to unsubstantiated messages, and determine whether the source of substantiation plays a role in persuasion.

Figure 2 evaluates the effect of different types of corrections on belief in misinformation, compared to the control condition (no correction). The first row of coefficients represents the size of the effect in corrections without any substantiation. In this case, a peer in the group chat expresses skepticism with the story but does not cite a source to justify their skepticism (merely stating a version of "I don't think that's true, bro!"). The remaining rows represent substantiated corrections, but each with a different source of substantiation. While the second series of estimates are effects for corrections in which the second user relied on a domain expert, the last row refers to corrections substanti-

ated by various fact-checking bodies.[13] In each case, we distinguish by story, including all seven stories that contained a misinformation stimulus.

Figure 2: The Effect of Different Types of Social Corrections, compared to Control Condition (Omitted Category)



Several striking findings emerge from these results. Visually, it appears that the corrective effect is larger on some stories when the correction is sourced; this appears particularly clear for the UNESCO and MMR rumors. Critically, however, potential differences in effect sizes remain, at best, very small across sub-types of corrections, and in most cases do not appear to exist at all, as confidence intervals between corrections largely overlap across rows. Given how highly powered our experiment is, we can say with relative confidence that the sophistication of social corrections mattered very little

[13]We further disentangle these results in Appendix L.

to respondents: the unsourced correction (a simple, short dismissal of the claim made by the first user) is often as effective as the longer and more clearly sourced corrections in this design. An unidentified participant merely expressing incredulity about a rumor is therefore often as likely to reduce belief in a falsehood as a respondent engaging in a longer correction.

We confirm these intuitions by comparing unsourced corrections to all other corrections and control (i.e. changing the omitted category in our model to unsourced corrections). We find that for 4 out of 7 stories, the effect of sourced corrections is not statistically distinguishable from that of unsourced corrections. That is, for the majority of rumors, unsourced corrections are just as effective as using sources. Moreover, in the few cases that sourced corrections do work better than unsubstantiated ones, their corrective effects remain small (see Appendix L).

The type of source cited appears to make even less of a difference: corrective messages substantiated with a domain expert do not make the correction more persuasive than other types of sourced corrections: in all cases, respondents are as likely to react to the correction when it is said to originate from a professional fact-checking organization, a prominent newspaper, or the platforms themselves, as opposed to a domain expert. This further implies that respondents open to belief change do not require much expertise in order for their beliefs to be moved.

Finally, when further disentangling these results by fact-checker (Appendix L), no source emerges as consistently more persuasive or effective relative to others. This may suggest that outsourcing fact-checking to credible authorities may not be necessary to improve its overall credibility in this context. On the contrary, merely indicating that a message might be false, relative to sourcing that indication, appears to be enough to move beliefs. Overall, we thus find that the content of social corrections counts less than the mere presence of one.

Discussion and Conclusion

Taken together, our results demonstrate that social corrections may be effective in contexts very different from the United States: in our experiment, exposure to a corrective message posted by an unidentified peer – regardless of the source or substantiation of that message – significantly reduces beliefs in misinformation. Insofar as respondents in our experiment were not incentivized to pay attention to the message, and by design neither knew nor by could identify the individuals posting corrections, these results may be seen as conservative estimates. Arguably, peer corrections on more homophilic networks (friends, colleagues, or likeminded partisans) may achieve a much larger effect.

While additional research will be needed to test this hypothesis, our results already point to a potential flip side of comparatively low levels of digital literacy. India has lower rates of formal education and digital literacy, as well as far more users who are new to the Internet. These factors likely imply that news received via the Internet might automatically have more value, given the unfamiliarity and fascination the medium inspires (Badrinathan 2021). While this may lead misinformation to be more easily believed in the first place, the same may apply to corrections, which may more easily become effective in such contexts.

Beyond our main result, tests on our other hypotheses (H2 to H5) point to another important reason why social corrections may play an important role in this context: forces such as partisan motivated reasoning, that are normally expected to reduce the effect of corrective measures, may not play as important a role in this context.

The absence of partisan motivated reasoning in our results - despite our multiple attempts at detecting any evidence of this well-researched trope - is striking. While we can only speculate as to the causes of this divergence, one explanation may lie in the nature of partisanship in India. Despite much report of political polarization and transformation of partisanship into a social identity (Chhibber and Verma 2018), India is a country that has traditionally had weaker partisan ties, and politics is thought to

be more clientelistic or ethnicity-based rather than programmatic (Auerbach et al. 2021). This relative weakness of partisanship - at least in the American sense of the term - may imply that motivated reasoning would *not* constitute as big an obstacle to correcting beliefs, or more likely in our view, that partisanship may not be the basis for motivated reasoning in this context (it may instead exist in another, non-partisan form, for instance, along the lines of religious identities).

Additionally, we show that respondents often do not react differently to substantiated and unsubstantiated corrections, and that the presence of a correction, rather than its degree of substantiation, appears sufficient to change beliefs. Both Munger (2017) and Siegel and Badaan (2020) show that online hatred and harassment can be significantly reduced by simple nudges, especially if these come from ingroups. Further,Groenendyk and Krupnikov (2020) demonstrate that information processing is motivated by the goals made salient in a given context. While political contexts may invoke conflict, WhatsApp groups formed around shared causes may push respondents to seek consensus around a common goal, making it easier for corrections that break the ice to have an effect.

While these multiple findings mutually reinforce our confidence that social corrections matter, we now outline limitations in our design pointing to the need for further research. First, we note that we measure our dependent variable in close proximity to the treatment, and thus cannot speak to the durability of these effects. Future studies should consider a longer gap between treatment and outcomes to measure whether such effects decay over time. Second, a challenging prospect for future research is to be able to examine the effect of corrections in a more naturalistic setting, outside of a survey or online experiment. While the private nature of WhatsApp groups makes this logistically and ethically difficult, measuring the impact of misinformation and solutions to counter it within the ecosystem of groups that individuals are a part of will allow us to ascertain the true impact of group solidarity, conformity pressures, and ingroup norms.

Even if they suggest that social corrections could become a useful tool in the

arsenal of measures to combat misinformation, the policy implications of our findings are mixed and complex. Our results imply that merely signaling a problem with the credibility of a rumor (regardless of *how* sophisticated this signaling is) may go a long way in reducing rates of beliefs in rumors. This may be seen as good news, as expecting users to post more sophisticated or substantiated corrections may be unrealistic: they may not know of fact-checking services; if they do, they may not be motivated to consult these services; if they did consult them and read their analyses, they may not be willing to invest time and energy in a lengthy explanation leading them to openly contradict one of their acquaintances; even if they were willing to take these steps, they may not find the right words.

A possible implication could be that users should be encouraged to effectively "sound off" as easily as possible to express their doubts about rumors posted on a chat. Platforms may help with this in a variety of ways. One way to reduce the cost of expressing doubt about a rumor may be to add a simple button to express doubt in reference to on-platform rumors, or enable users to easily flag statements as problematic, unreliable, or groundless. Such a strategy would be entirely compatible with the encrypted nature of the platform, as red flags need not be reported or investigated by the platform, but merely used to communicate to other users. Such a strategy would, in addition, allow a single user to very quickly flag a large number of posts, and hence more effectively combat the barrage of misinformation that currently exists on these platforms.

Yet, encouraging users to sound off as easily as possible may have perverse consequences, especially if users are easily influenced. True stories could be perniciously "corrected", and our results suggest that such misplaced "corrections" may likely be believed.[14] This may point to the need for combining social corrections (and corrections, more generally) with other strategies, such as digital literacy or inoculation theory training. In order for users to efficiently correct one another, they must be trained to

[14]While this raises ethical challenges, future research could analyze whether a peer posting a factually incorrect correction may have a similar effect on belief change.

recognize credible corrections from manipulative ones.

Beyond chat apps and other messaging applications, our study opens up broader avenues for research on misinformation in developing countries. Much remains to be uncovered about the ability of misinformation to persuade, and to be corrected, in settings of low education, accelerating Internet, and private online spaces. The weakness of the partisan form of motivated reasoning detected in our study suggests that more comparative work on misinformation is needed. Future work should explore the psychological mechanisms leading to belief change, and potentially to offline behaviors, especially in countries where the stakes are as high as violence. Such research should also look into information and misinformation processing on encrypted and personal social media networks such as WhatsApp. The findings from this study have implications not only for developing countries that widely use EMAs, but also for more developed contexts where polarized users are sorted into homophilic networks online.

References

Arun, Chinmayi. 2019. "On WhatsApp, Rumours, Lynchings, and the Indian Government." *Economic & Political Weekly* 54 (6).

Auerbach, Adam Michael, Jennifer Bussell, Simon Chauchard, Francesca R. Jensenius, Gareth Nellis, Mark Schneider, Neelanjan Sircar, Pavithra Suryanarayan, Tariq Thachil, Milan Vaishnav, and et al. 2021. "Rethinking the Study of Electoral Politics in the Developing World: Reflections on the Indian Case." *Perspectives on Politics*: 1–15.

Badrinathan, Sumitra. 2021. "Educative Interventions to Combat Misinformation: Evidence From A Field Experiment in India." *American Political Science Review*: 1-17.

Bode, Leticia, and Emily K Vraga. 2018. "See Something, Say Something: Correction of Global Health Misinformation on Social Media." *Health Communication* 33 (9): 1131–1140.

Bussell, Jennifer. 2019. *Clients and Constituents: Political Responsiveness in Patronage Democracies*. New York: Oxford University Press.

Chan, Man-pui Sally, Christopher R Jones, Kathleen Hall Jamieson, and Dolores Albarracín. 2017. "Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation." *Psychological Science* 28 (11): 1531–1546.

Chandra, Kanchan. 2004. *Why Ethnic Parties Succeed: Patronage and Ethnic Headcounts in India*. New York: Cambridge University Press.

Chhibber, Pradeep K, and Rahul Verma. 2018. *Ideology and Identity: The Changing Party Systems of India*. New York: Oxford University Press.

Eagly, Alice H, and Shelly Chaiken. 1993. *The psychology of attitudes.* Harcourt Brace Jovanovich College Publishers.

Farooq, Gowhar. 2017. "Politics of Fake News: how WhatsApp became a potent propaganda tool in India." *Media Watch* 9 (1): 106–117.

Flynn, DJ, Brendan Nyhan, and Jason Reifler. 2017. "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics." *Political Psychology* 38: 127–150.

Fridkin, Kim, Patrick J Kenney, and Amanda Wintersieck. 2015. "Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising." *Political Communication* 32 (1): 127–151.

Garimella, Kiran, and Dean Eckles. 2020. "Images and misinformation in political groups: evidence from WhatsApp in India." *Harvard Kennedy School Misinformation Review*.

German, Kathleen M, Bruce E Gronbeck, Douglas Ehninger, and Alan H Monroe. 2016. *Principles of public speaking*. New York: Routledge.

Green, Donald P, Bradley Palmquist, and Eric Schickler. 2004. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. New Haven: Yale University Press.

Groenendyk, Eric, and Yanna Krupnikov. 2020. "What Motivates Reasoning? A Theory of Goal-Dependent Political Evaluation." *American Journal of Political Science*: 1–17.

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan 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): 15536–15545.

Housholder, Elizabeth E, and Heather L LaMarre. 2014. "Facebook Politics: Toward a Process Model for Achieving Political Source Credibility Through Social Media." *Journal of Information Technology & Politics* 11 (4): 368–382.

Martel, Cameron, Mohsen Mosleh, and David Gertler Rand. 2021. "You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online." *Media and Communication*: 120-133.

Michelitch, Kristin, and Stephen Utych. 2018. "Electoral Cycle Fluctuations in Partisanship: Global Evidence from Eighty-Six Countries." *The Journal of Politics* 80 (2): 412–427.

Munger, Kevin. 2017. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39 (3): 629–649.

Neyazi, Taberez Ahmed, Antonis Kalogeropoulos, and Rasmus K Nielsen. 2021. "Misinformation Concerns and Online News Participation among internet Users in India." *Social Media+ Society* 7 (2): 20563051211009013.

Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32 (2): 303–330.

Pereira, Frederico Batista, Natália S Bueno, Felipe Nunes, and Nara Pavão. 2021. "Motivated Reasoning Without Partisanship? Fake News in the 2018 Brazilian Elections." *Working Paper*.

Porter, Ethan, and Thomas J Wood. 2021. "The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom." *Proceedings of the National Academy of Sciences* 118 (37).

Reinard, John C. 1988. "The Empirical Study of the Persuasive Effects of Evidence The Status After Fifty Years of Research." *Human Communication Research* 15 (1): 3–59.

Rosenzweig, Leah R, Bence Bago, Adam J Berinsky, and David G Rand. 2021. "Happiness and surprise are associated with worse truth discernment of COVID-19

headlines among social media users in Nigeria." *Harvard Kennedy School Misinformation Review*.

Rossini, Patrícia, Jennifer Stromer-Galley, Erica Anita Baptista, and Vanessa Veiga de Oliveira. 2020. "Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections." *New Media & Society*: 14-29.

Siegel, Alexandra A, and Vivienne Badaan. 2020. "# No2Sectarianism: Experimental approaches to reducing sectarian hate speech online." *American Political Science Review* 114 (3): 837–855.

Stiff, James B, and Paul A Mongeau. 2016. *Persuasive Communication*. New York: Guilford Publications.

Swire, Briony, and Ullrich K H Ecker. 2018. "Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication." *Misinformation and Mass Audiences*: 195–211.

Taber, Charles S, and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–769.

Thorson, Emily. 2016. "Belief Echoes: The Persistent Effects of Corrected Misinformation." *Political Communication* 33 (3): 460–480.

Tokita, Christopher K, Andrew M Guess, and Corina E Tarnita. 2021. "Polarized information ecosystems can reorganize social networks via information cascades." *Proceedings of the National Academy of Sciences* 118 (50).

van der Meer, Toni GLA, and Yan Jin. 2020. "Seeking Formula for Misinformation Treatment in Public Health Crises: The Effects of Corrective Information Type and Source." *Health Communication* 35 (5): 560–575.

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

Vraga, Emily K, and Leticia Bode. 2018. "I do not believe you: how providing a source corrects health misperceptions across social media platforms." *Information, Communication & Society* 21 (10): 1337–1353.

Supplementary Information File

"I Don't Think That's True, Bro!"

Social Corrections for Misinformation in India

Contents

A 2019 WhatsApp Campaign Promoting User-driven Corrections

Figure A.1: WhatsApp ad in Indian newspapers

B Advertisement Used to Recruit Respondents

The image below is a screenshot of the advertisement we used to recruit respondents into our survey via Facebook.

Figure B.1: Sample recruitment ad

C Full Text of Experimental Manipulations

The text displayed on our WhatsApp screenshots to treatment group respondents varied as a function of the type of rumor, the politician / media outlet it is attributed to, and the nature of the correction. Figure C.1 below describes the full text associated with each hypothetical WhatsApp screenshot in each condition. Figure C.2 shows an example of the treatment stimuli.

		SOURCES VARIATIONS				CORRECTION VARIATIONS			
Rumor	Type of claim	Message 1	Message 2 (part 1)	Message 2 (part 2)		Message 3a (control)	Message 3b (Domain	M3b (Factchecker correction)	M3c (random person
1	Congenial to BJP/dissonant to non-BJP	image	Check this out.	In this article published on the NDTV Hindi/Doordarshan/India TV website, a BJP leader claims that the Muslim population in India will overtake the Hindu population in the future.	In this article published on the NDTV Hindi/Doordarshan/India TV website, it is suggested that the Muslim population in India will overtake the Hindu population in India in the future.	thanks -- will read!	The other day I saw a report by prominent demographers that said this was extremely unlikely to happen. So I don't think that's right actually.	The other day I saw a fact-checking report by ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp that said this was extremely unlikely to happen. So I don't think that's right actually.	The other day I saw something that said this was extremely unlikely to happen. So I don't think that's right actually.
2	Congenial to BJP/dissonant to non-BJP	image	Just saw this!	This appeared on the NDTV Hindi/Doordarshan/India TV website --- A BJP leader explains that polygamy (having several wifes) is very common in the Muslim population.	This appeared on the NDTV Hindi/Doordarshan/India TV website --- it says that polygamy (having several wifes) is very common in the Muslim population.	Thanks. Will check when I have a second.	Hmm, not sure. The Census of India did a report and showed it to be unsubstantiated, though.	Hmm, not sure. A fact-checking report done by ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp showed this to be unsubstantiated, though.	Hmm, not sure. I saw something showed this to be unsubstantiated, though.
3	Neither congenial nor dissonant to BJP/non-BJP	image	Just came across this article	-	This comes from the NDTV Hindi/Doordarshan/India TV website. Apparently M-R vaccines are associated with autism and retardation.	Wow, ok. will get into this.	Hey I don't think that's true actually. I just saw a report from doctors from AIIMS, there appears to be no basis for this claim...	Hey I don't think that's true actually. I just saw a fact-checking report done by ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp, there appears to be no basis for this claim...	Hey I don't think that's true actually. Someone told me there was not basis for this claim...
4	Congenial to BJP/dissonant to non-BJP	image	This is worth looking at.	The NDTV Hindi/Doordarshan/India TV website just published this. A bunch of BJP leaders said that drinking cow urine (gomutra) helps build one's immune system.	The NDTV/NDTV Hindi/Doordarshan/RepublicTV/India TV website just published this. Claims that Drinking cow urine (gomutra) helps build one's immune system.	Got it, thanks for sending :)	Actually not sure about this, brother. I saw a report from doctors from AIIMS explaining why this is not correct.	Actually not sure about this, brother. I saw a fact-checking report done by ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp explaining why this is not correct.	Actually not sure about this, brother. I saw somewhere that this is not correct.
5	Neither congenial nor dissonant to BJP/non-BJP	image	Relevant as the ICC world cup approaches...	-	This comes from the NDTV Hindi/Doordarshan/India TV website. I had forgotten that Australia has more ICC cricket world cup wins than any country!	Great. Thanks for sending :)	-	-	-
6	Neither congenial nor dissonant to BJP/non-BJP	image	Important stuff...	-	the NDTV/NDTV Hindi/Doordarshan/RepublicTV/India TV website published this. sad that there's still no cure for HIV/AIDS	thanks. will definitely read.	-	-	-
7	Congenial to non-BJP/dissonant to BJP	image	Just saw this!	NDTV Hindi/Doordarshan/India TV: several INC leaders claim that the BJP hacks electronic voting machines.	NDTV Hindi/Doordarshan/India TV: some people suggesting that the BJP hacks electronic voting machines.	ok! reading now...	Not sure about this... the Election Commission released a serious report saying there's no basis for this claim	Not sure about this claim. ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp has come up with a detailed fact-checking report that showed there was no basis for this argument.	Not sure about this claim. I saw somewhere there is no basis for this argument.
8	Congenial to BJP/dissonant to non-BJP	image	Wow	Just saw this on the NDTV Hindi/Doordarshan/India TV weblsite.. This BJP guy said UNESCO declared PM Modi best Prime Minister in 2016.	Just saw this on NDTV Hindi/Doordarshan/India TV website.. UNESCO declared PM Modi best Prime Minister in 2016!	Thanks, boss :)	Haha that's not right actually.. UNESCO put out a release saying they didn't come up with rankings like that.	Haha that's not right actually.. ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp published a fact-checking thing saying that UNESCO didn't come up with rankings like that.	Haha that's not right actually..
9	Neither congenial nor dissonant to BJP/non-BJP	image	Have a look at this!	From the NDTV Hindi/Doordarshan/India TV website ... Netaji Bose did NOT die in a plane crash in 1945!		wow - thanks for sharing!	This theory has been debunked, I think. I read a report by Delhi University historians explaining there was no ground to believe any of this.	This theory has been debunked, I think. I read a fact-checking report by ALTNEWS/Vishwasnews.com/Times of India/Facebook/WhatsApp explaining there was no ground to believe any of this.	I think this theory has been debunked, though.

Figure C.1: Text for experimental manipulations

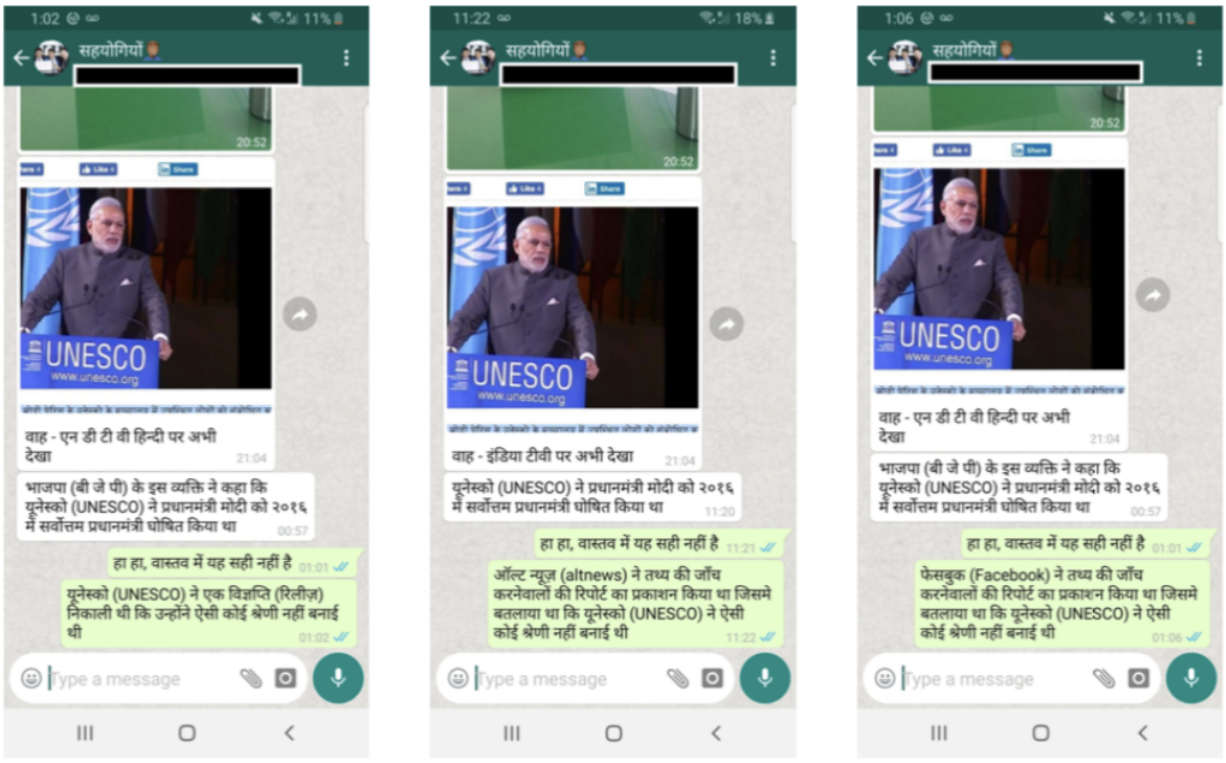


Figure C.2: Sample Treatment Stimuli

D Tests For Hypothesis 2

Hypothesis 2a: Peer corrections will be more effective when misinformation is attributed to an ideologically dissonant politician (compared to when it is unattributed).

To test this hypothesis, we run the following model:

$$\begin{aligned} Belief\,Accuracy_i = \alpha + \beta_1 AnyCorrection_i + \beta_2 DissonantPol_i \\ + \beta_3 AnyCorrection * DissonantPol_i + \epsilon_i \end{aligned} \tag{D.1}$$

As noted in the body of the article, we limit our analyses to the subset of rumors that are clearly partisan in nature (rumors 3, 4, 6, 8, and 9) and code whether the claim was attributed in the prompt to a congenial or dissonant politician. We code a politician as congenial or dissonant as a function of the respondent's partisan inclination towards the BJP (the ruling party), relying on the respondent's expressed closeness to this party. A BJP politician is deemed congenial if the respondent describes herself as close or very close to the party and dissonant if the respondent describes herself as far or very far from the party. By contrast, a INC politician is deemed congenial if the respondent describes herself as far or very far to the BJP and dissonant if the respondent describes herself as close or very close to the BJP. Note that we are pooling members of both major parties in each category (e.g., "dissonant" takes the value of 1 for BJP identifiers who read an anti-BJP claim and for INC identifiers who read a pro-BJP claim).

Table D.1: Effect of Any Correction * Dissonant Speaker on Belief in Rumor

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
AnyCorrection	-0.127^{***}	-0.178^{***}	-0.092^{***}	-0.007	-0.405^{***}
	(0.038)	(0.031)	(0.033)	(0.033)	(0.040)
DissonantPol	-0.564^{***}	-0.310^{*}	-0.242	-0.229^{**}	-0.184
	(0.149)	(0.164)	(0.154)	(0.108)	(0.203)
AnyCorrection*	0.482^{***}	0.062	-0.055	0.008	0.016
DissonantPol	(0.163)	(0.174)	(0.168)	(0.118)	(0.218)
Constant	2.775^{***}	3.280^{***}	3.018^{***}	1.666^{***}	3.005^{***}
	(0.034)	(0.026)	(0.028)	(0.028)	(0.034)
Observations	5,104	5,103	5,099	5,136	5,109
R^2	0.005	0.012	0.006	0.005	0.022
Adjusted R^2	0.004	0.011	0.006	0.005	0.021
Res. Std. Er.	1.093 (df = 5100)	0.944 (df = 5099)	1.058 (df = 5095)	1.011 (df = 5132)	1.250 (df = 5105)
F Statistic	7.934^{***}	20.647^{***}	10.792^{***}	9.195^{***}	37.700^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

Hypothesis 2b: Peer corrections will be less effective when misinformation is attributed to an ideologically congenial politician (compared to when it is unattributed).

$$BeliefAccuracy_i = \alpha + \beta_1 AnyCorrection_i + \beta_2 CongenialPol_i + \beta_3 AnyCorrection * CongenialPol_i + \epsilon_i \quad \text{(D.2)}$$

Table D.2: Effect of Any Correction * Congenial Speaker on Belief in Rumor

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
AnyCorrection	-0.102^{***}	-0.181^{***}	-0.131^{***}	-0.043	-0.400^{***}
	(0.039)	(0.031)	(0.034)	(0.032)	(0.041)
CongenialPol	0.070	0.167	-0.011	0.518^{***}	0.409^{***}
	(0.132)	(0.107)	(0.119)	(0.180)	(0.157)
AnyCorrection*	-0.054	-0.155	0.176	-0.120	-0.349^{**}
CongenialPol	(0.141)	(0.116)	(0.129)	(0.191)	(0.167)
Constant	2.741^{***}	3.262^{***}	3.011^{***}	1.638^{***}	2.979^{***}
	(0.034)	(0.027)	(0.028)	(0.027)	(0.035)
Observations	5,104	5,103	5,099	5,136	5,109
R^2	0.002	0.008	0.004	0.010	0.022
Adjusted R^2	0.001	0.008	0.004	0.009	0.022
Res. Std. Error	1.095 (df = 5100)	0.946 (df = 5099)	1.059 (df = 5095)	1.009 (df = 5132)	1.250 (df = 5105)
F Statistic	2.747^{**}	14.136^{***}	7.207^{***}	16.946^{***}	38.594^{***}
Note:					*p<0.1; **p<0.05; ***p<0.01

E Tests For Hypothesis 3

Hypothesis 3a: Peer corrections will be more effective when misinformation originates from a dissonant media outlet (compared to unattributed or neutral outlet).

Hypothesis 3b: Peer corrections will be less effective when misinformation originates from a congenial media outlet (compared to an unattributed or neutral outlet).

To test these hypotheses, we code a media outlet as congenial or dissonant as a function of the respondent's expressed proximity to the BJP. Concretely, we code the "pro-BJP" outlet (here, India TV) as congenial and the "anti-BJP" outlet (here, New Delhi TV or NDTV) as dissonant when the respondent reports feeling close or very close to the BJP. By contrast, we code the "pro-BJP" outlet (India TV) as dissonant and "anti-BJP" outlet (NDTV) as congenial when the respondent reports feeling far or very far to the BJP.

We test this hypothesis with the following model:

$$\begin{aligned} BeliefAccuracy_i = \alpha + \beta_1 AnyCorrection_i + \beta_2 CongenialMedia_i + \beta_3 DissonantMedia_i + \\ \beta_4 AnyCorrection * CongenialMedia_i + \beta_5 AnyCorrection * DissonantMedia_i + \epsilon_i \end{aligned} \tag{E.1}$$

Table E.1: Effect of Any Correction * Media Outlet Source on Belief in Rumor

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AnyCorrection	-0.150^{***}	-0.167^{***}	-0.419^{***}	-0.128^{***}	-0.006	-0.399^{***}	-0.110^{***}
	(0.041)	(0.033)	(0.036)	(0.036)	(0.035)	(0.043)	(0.034)
Congenial Media	-0.224^{*}	0.230^{**}	-0.078	0.080	-0.163	0.036	0.186
	(0.126)	(0.113)	(0.121)	(0.108)	(0.111)	(0.151)	(0.114)
Dissonant Media	-0.132	0.061	-0.050	-0.184	0.051	0.094	-0.048
	(0.121)	(0.108)	(0.121)	(0.116)	(0.118)	(0.143)	(0.116)
AnyCorrection* CongenialMedia	0.271^{**}	-0.194	-0.049	-0.056	0.080	-0.067	-0.174
	(0.135)	(0.122)	(0.131)	(0.119)	(0.121)	(0.162)	(0.125)
AnyCorrection* DissonantMedia	0.219^{*}	-0.134	-0.067	0.272^{**}	-0.134	-0.092	0.091
	(0.131)	(0.116)	(0.130)	(0.126)	(0.127)	(0.155)	(0.127)
Constant	2.773^{***}	3.256^{***}	2.834^{***}	3.015^{***}	1.658^{***}	2.992^{***}	2.781^{***}
	(0.036)	(0.027)	(0.030)	(0.029)	(0.029)	(0.036)	(0.027)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.003	0.009	0.038	0.003	0.002	0.021	0.003
Adjusted R^2	0.002	0.008	0.037	0.002	0.001	0.020	0.002
Res. Std. Er.	1.095	0.945	1.038	1.060	1.013	1.251	1.048
F Statistic	3.055^{***}	9.650^{***}	39.456^{***}	3.394^{***}	1.663	21.697^{***}	3.190^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

F Heterogeneous Effects of BJP Support

To complement our tests of motivated reasoning (based on the congeniality/dissonance of the information presented and the source of the information), we present OLS results from models that test whether BJP voters react differently to corrective information.

$$\begin{aligned} Belief\,Accuracy_i = \alpha + \beta_1 AnyCorrection_i + \beta_2 BJPSupport_i \\ + \beta_3 AnyCorrection * BJPSupport_i + \epsilon_i \end{aligned} \tag{F.1}$$

Table F.1: Effect of BJP Support * Correction

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
AnyCorrection	−0.078	−0.168***	−0.452***	−0.027	−0.123**	−0.405***	−0.058
	(0.064)	(0.053)	(0.058)	(0.056)	(0.052)	(0.068)	(0.054)
BJP Support	0.424***	0.338***	0.092	0.578***	−0.870***	0.491***	0.237***
	(0.069)	(0.055)	(0.060)	(0.057)	(0.054)	(0.071)	(0.053)
AnyCorrection *	−0.043	−0.027	0.006	−0.114*	0.151**	−0.017	−0.081
BJP Support	(0.078)	(0.064)	(0.070)	(0.068)	(0.064)	(0.083)	(0.066)
Constant	2.460***	3.040***	2.765***	2.616***	2.238***	2.669***	2.629***
	(0.057)	(0.045)	(0.050)	(0.047)	(0.045)	(0.058)	(0.044)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.029	0.032	0.037	0.051	0.124	0.052	0.009
Adjusted R^2	0.029	0.032	0.037	0.050	0.123	0.051	0.009
Res. Std. Er.	1.080	0.934	1.038	1.034	0.949	1.231	1.045
F Statistic	51.459***	56.702***	65.187***	90.418***	241.189***	93.117***	16.129***
Note:							*p<0.1; **p<0.05; ***p<0.01

G Main Effect of Congenial / Dissonant Claim

In this section, we show that the claims we code as congenial to respondents are more likely to be believed (G.1) and that the claims we code as dissonant to respondents are less likely to be believed (G.2). In each case we run a simple bivariate OLS model:

$$Belief = \alpha + \beta_1(CongenialClaim/DissonantClaim) + \epsilon \tag{G.1}$$

Table G.1: Effect of Rumor Congeniality on Belief

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
CongenialClaim	0.218***	0.214***	0.378***	0.480***	0.309***
	(0.031)	(0.027)	(0.030)	(0.029)	(0.036)
Constant	2.530***	3.001***	2.702***	1.478***	2.506***
	(0.025)	(0.021)	(0.024)	(0.017)	(0.028)
Observations	5,104	5,103	5,099	5,136	5,109
R^2	0.009	0.012	0.030	0.049	0.014
Adjusted R^2	0.009	0.012	0.030	0.049	0.014
Res. Std. Er.	1.091 (df = 5102)	0.944 (df = 5101)	1.045 (df = 5097)	0.989 (df = 5134)	1.255 (df = 5107)
F Statistic	48.141***	62.228***	157.494***	264.374***	73.238***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table G.2: Effect of Rumor Dissonance on Belief

	Dependent variable: Belief in Rumor				
	MuslimPop	Polygamy	Gomutra	EVM	UNESCO
	(1)	(2)	(3)	(4)	(5)
DissonantClaim	−0.157***	−0.176***	−0.309***	−0.625***	−0.231***
	(0.033)	(0.028)	(0.031)	(0.028)	(0.038)
Constant	2.716***	3.190***	3.035***	2.019***	2.772***
	(0.019)	(0.016)	(0.018)	(0.022)	(0.021)
Observations	5,104	5,103	5,099	5,136	5,109
R^2	0.004	0.008	0.019	0.090	0.007
Adjusted R^2	0.004	0.007	0.018	0.089	0.007
Res. Std. Er.	1.093 (df = 5102)	0.946 (df = 5101)	1.051 (df = 5097)	0.967 (df = 5134)	1.259 (df = 5107)
F Statistic	22.998***	38.607***	96.136***	505.256***	37.676***
Note:					*p<0.1; **p<0.05; ***p<0.01

H Summary Statistics

Table H.1: Summary Statistics for Muslim Population Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,104	2.665	1.096	1	3	4
Any Correction	5,104	0.781	0.414	0	1	1
Outpartisan Speaker	5,104	0.069	0.253	0	0	1
Copartisan Speaker	5,104	0.126	0.332	0	0	1
Congenial Media	5,104	0.134	0.341	0	0	1
Dissonant Media	5,104	0.127	0.333	0	0	1
Congenial Claim	5,104	0.616	0.486	0	1	1
Dissonant Claim	5,104	0.326	0.469	0	0	1
BJP Partisan	5,104	0.678	0.467	0	1	1
Congress Partisan	5,104	0.057	0.233	0	0	1
Pure Control	5,104	0.023	0.148	0	0	1
Peer Correction	5,104	0.258	0.438	0	0	1
Expert Correction	5,104	0.261	0.439	0	0	1
Alt News	5,104	0.051	0.220	0	0	1
Vishwas	5,104	0.053	0.225	0	0	1
TOI	5,104	0.048	0.213	0	0	1
Facebook	5,104	0.054	0.226	0	0	1
WhatsApp	5,104	0.056	0.229	0	0	1

Table H.2: Summary Statistics for Polygamy Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,103	3.133	0.949	1	3	4
Any Correction	5,103	0.735	0.441	0	1	1
Outpartisan Speaker	5,103	0.063	0.243	0	0	1
Copartisan Speaker	5,103	0.118	0.323	0	0	1
Congenial Media	5,103	0.116	0.321	0	0	1
Dissonant Media	5,103	0.127	0.333	0	0	1
Congenial Claim	5,103	0.618	0.486	0	1	1
Dissonant Claim	5,103	0.323	0.468	0	0	1
BJP Partisan	5,103	0.680	0.467	0	1	1
Congress Partisan	5,103	0.058	0.234	0	0	1
Pure Control	5,103	0.022	0.145	0	0	1
Peer Correction	5,103	0.247	0.431	0	0	1
Expert Correction	5,103	0.247	0.431	0	0	1
AltNews	5,103	0.045	0.208	0	0	1
Vishwas	5,103	0.050	0.219	0	0	1
TOI	5,103	0.048	0.215	0	0	1
Facebook	5,103	0.047	0.212	0	0	1
WhatsApp	5,103	0.050	0.218	0	0	1

Table H.3: Summary Statistics for MMR Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,061	2.500	1.058	1	3	4
Any Correction	5,061	0.729	0.444	0	1	1
Congenial Media	5,061	0.125	0.331	0	0	1
Dissonant Media	5,061	0.121	0.326	0	0	1
BJP Partisan	5,061	0.680	0.466	0	1	1
Congress Partisan	5,061	0.058	0.234	0	0	1
Pure Control	5,061	0.022	0.146	0	0	1
Peer Correction	5,061	0.234	0.424	0	0	1
Expert Correction	5,061	0.243	0.429	0	0	1
AltNews	5,061	0.054	0.227	0	0	1
Vishwas	5,061	0.053	0.224	0	0	1
TOI	5,061	0.050	0.218	0	0	1
Facebook	5,061	0.046	0.210	0	0	1
WhatsApp	5,061	0.049	0.215	0	0	1

Table H.4: Summary Statistics for Gomutra Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,099	2.935	1.061	1	3	4
Any Correction	5,099	0.706	0.455	0	1	1
Outpartisan Speaker	5,099	0.061	0.240	0	0	1
Copartisan Speaker	5,099	0.121	0.326	0	0	1
Congenial Media	5,099	0.128	0.334	0	0	1
Dissonant Media	5,099	0.127	0.334	0	0	1
Congenial Claim	5,099	0.618	0.486	0	1	1
Dissonant Claim	5,099	0.323	0.468	0	0	1
BJP Partisan	5,099	0.680	0.467	0	1	1
Congress Partisan	5,099	0.058	0.233	0	0	1
Pure Control	5,099	0.023	0.151	0	0	1
Peer Correction	5,099	0.204	0.403	0	0	1
Expert Correction	5,099	0.251	0.433	0	0	1
AltNews	5,099	0.053	0.224	0	0	1
Vishwas	5,099	0.051	0.221	0	0	1
TOI	5,099	0.048	0.214	0	0	1
Facebook	5,099	0.052	0.222	0	0	1
WhatsApp	5,099	0.046	0.209	0	0	1

Table H.5: Summary Statistics for EVM Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,136	1.633	1.014	1	1	4
Any Correction	5,136	0.728	0.445	0	1	1
Outpartisan Speaker	5,136	0.125	0.331	0	0	1
Copartisan Speaker	5,136	0.063	0.243	0	0	1
Congenial Media	5,136	0.125	0.331	0	0	1
Dissonant Media	5,136	0.125	0.331	0	0	1
Congenial Claim	5,136	0.323	0.468	0	0	1
Dissonant Claim	5,136	0.618	0.486	0	1	1
Congress Partisan	5,136	0.057	0.232	0	0	1
BJP Partisan	5,136	0.680	0.467	0	1	1
Pure Control	5,136	0.022	0.148	0	0	1
Peer Correction	5,136	0.250	0.433	0	0	1
Expert Correction	5,136	0.221	0.415	0	0	1
AltNews	5,136	0.051	0.220	0	0	1
Vishwas	5,136	0.053	0.224	0	0	1
TOI	5,136	0.051	0.220	0	0	1
Facebook	5,136	0.055	0.227	0	0	1
WhatsApp	5,136	0.048	0.214	0	0	1

Table H.6: Summary Statistics for UNESCO Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,109	2.697	1.264	1	3	4
Any Correction	5,109	0.734	0.442	0	1	1
Outpartisan Speaker	5,109	0.059	0.235	0	0	1
Copartisan Speaker	5,109	0.118	0.322	0	0	1
Congenial Media	5,109	0.119	0.324	0	0	1
Dissonant Media	5,109	0.119	0.323	0	0	1
Congenial Claim	5,109	0.619	0.486	0	1	1
Dissonant Claim	5,109	0.324	0.468	0	0	1
Congress Partisan	5,109	0.057	0.233	0	0	1
BJP Partisan	5,109	0.681	0.466	0	1	1
Pure Control	5,109	0.010	0.099	0	0	1
Peer Correction	5,109	0.253	0.435	0	0	1
Expert Correction	5,109	0.249	0.433	0	0	1
AltNews	5,109	0.049	0.215	0	0	1
Vishwas	5,109	0.044	0.204	0	0	1
TOI	5,109	0.050	0.218	0	0	1
Facebook	5,109	0.047	0.211	0	0	1
WhatsApp	5,109	0.042	0.202	0	0	1

Table H.7: Summary Statistics for Bose Rumor

Statistic	N	Mean	St. Dev.	Min	Median	Max
Belief in Rumor	5,117	2.716	1.049	1	3	4
Any Correction	5,117	0.663	0.473	0	1	1
Congenial Media	5,117	0.126	0.331	0	0	1
Dissonant Media	5,117	0.114	0.317	0	0	1
BJP Partisan	5,117	0.681	0.466	0	1	1
Congress Partisan	5,117	0.057	0.232	0	0	1
Pure Control	5,117	0.023	0.149	0	0	1
Peer Correction	5,117	0.252	0.434	0	0	1
Expert Correction	5,117	0.175	0.380	0	0	1
AltNews	5,117	0.047	0.211	0	0	1
Vishwas	5,117	0.045	0.207	0	0	1
TOI	5,117	0.047	0.212	0	0	1
Facebook	5,117	0.048	0.214	0	0	1
WhatsApp	5,117	0.048	0.214	0	0	1

I Pretest Data

We ran a pretest on a panel of Facebook-recruited Indian respondents in early May 2019 (N=640) to measure the salience and rate of belief in 37 different rumors commonly disseminated on social media in India. These rumors were:

1. In the future, the Muslim population in India will overtake the Hindu population in India.
2. Polygamy is very common in the Muslim population.
3. Papaya leaf juice is a good way to cure dengue fever.
4. The food prepared by menstruating women is contaminated and rots faster.
5. M-R vaccines are associated with autism and retardation.
6. M-R vaccines are sometimes used by the government to control the population growth amongst certain groups.
7. One must sleep on the left side after having food, as any other sleeping position could be harmful to the digestive tract.
8. Drinking cow urine (gomutra) can help build one's immune system.
9. Gandhi did not try to save Baghat Singh and may even have been a co-conspirator in his death.
10. Indira Gandhi converted to Islam after marrying Feroze Gandhi.
11. Netaji Bose did NOT die in a plane crash in 1945.
12. Arvind Kejriwal has a drinking problem and makes videos while drunk.
13. Sonia Gandhi smuggled Indian treasures to Italy.
14. The BJP has hacked electronic voting machines.
15. NRIs will be able to vote online during the 2019 elections.
16. New Indian notes have a GPS chip to detect black money.
17. UNESCO declared PM Modi best Prime Minister in 2016.
18. WhatsApp profile pictures can be used by ISIS for terror activities.

19. People with cancer shouldn't eat sugar as it feeds cancer cells.

20. Biopsy causes a tumour to turn cancerous.

21. One should not take the P/500 paracetamol, as doctors have shown it to contain machupo, one of the most dangerous viruses in the world.

22. Dengue can be prevented with coconut oil, cardamom seeds, and eupatorium perfoliatum.

23. Amul Kulfi has some pig contents.

24. Drinking Pepsi after eating Polo or Mentos can cause instant death.

25. The BJP is in league with Facebook to remove anti-BJP pages and advertisements.

26. PM Modi hired a makeup artist for 15 lakh monthly salary.

27. Amit Shah personally ordered the assassination of Judge Loya.

28. Arun Jaitley is the current minister of Finance of the Government of India.

29. Scientists warn that current air quality in Delhi shortens lifespan by several years on average.

30. Pryanka Chopra married an American singer in 2018.

31. Mukesh Ambani's residence in Mumbai is the largest private home in the world.

32. India is now the fifth largest economy in the world.

33. Sachin Tendulkar owns the record number of runs record in the ICC cricket world cup.

34. Australia is the country that has won the ICC cricket world cup the most often.

35. According to the 2011 census, Sikhs represent less than 2% of the total Indian population.

36. There is no vaccine that cures HIV/AIDS.

37. Gandhi started his political career in South Africa before coming back to India.

In Figure I.1 we plot we percent of the pretest sample who said they heard each rumor. In Figure I.2 we plot the percent of the sample who said a given rumor was very accurate or somewhat accurate. We highlight the rumors from this list that we selected for the final experiment.

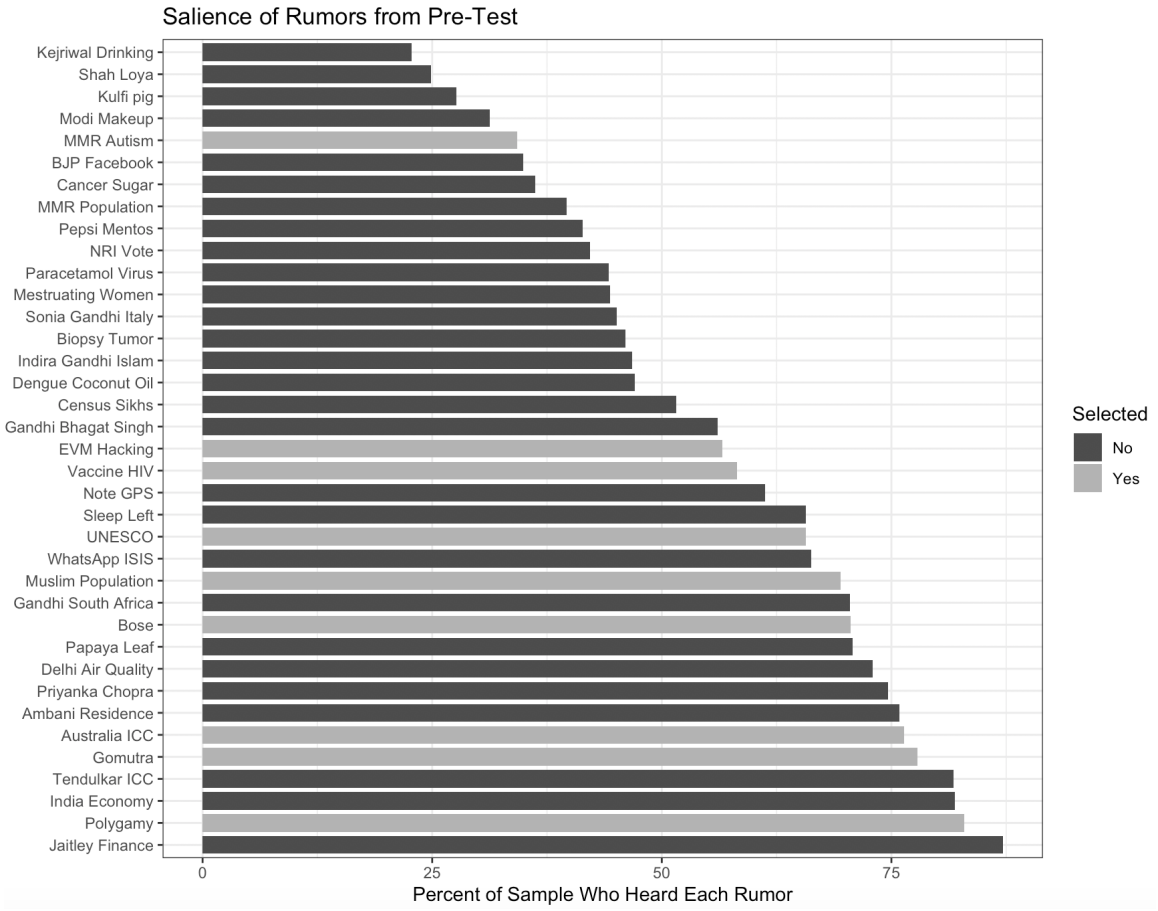


Figure I.1: Salience of Pretest Rumors

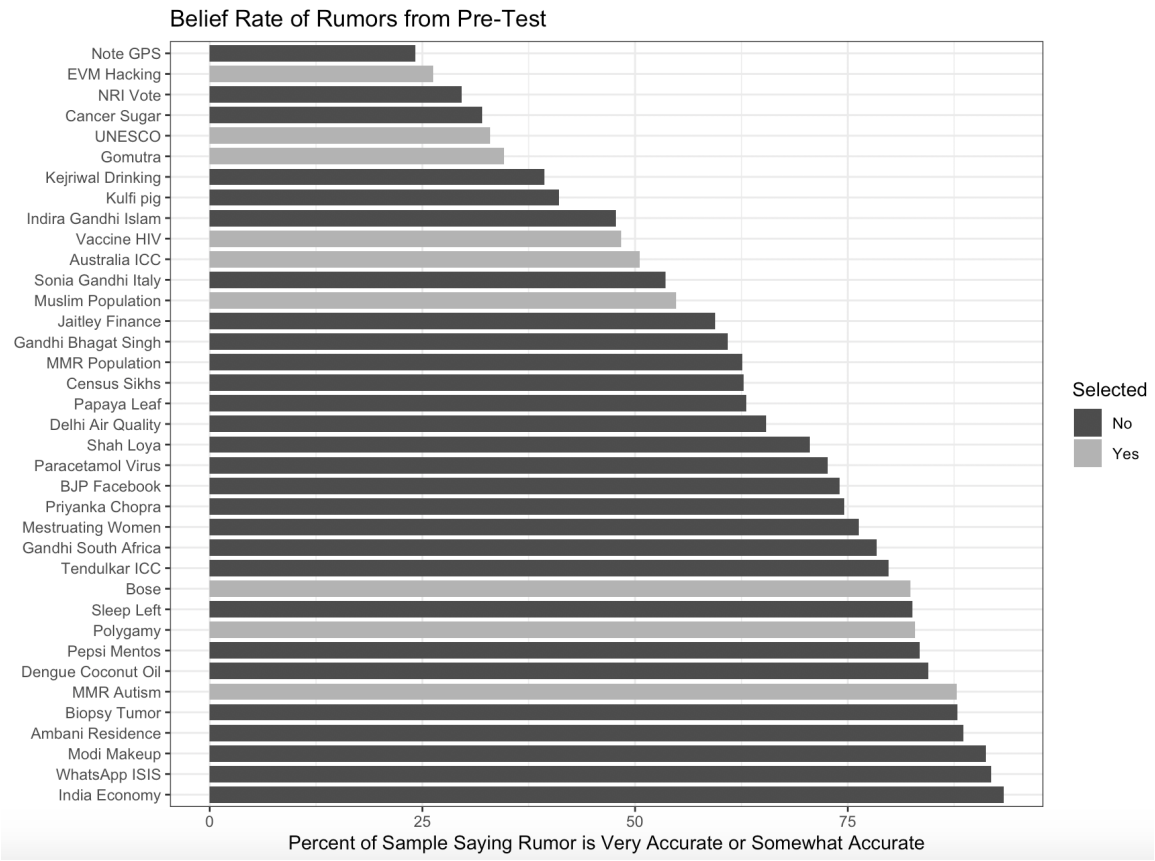


Figure I.2: Belief in Pretest Rumors

J Comparing the Effect of Control vs. Pure Control on Belief in Rumors

In this section, we restrict our sample to items for which respondents received either the control condition ("neutral" reaction by a second user but no correction) or the pure control (no screenshot; respondents directly asked the dependent variable). In the regressions presented below, we test in this sub-sample the effect of receiving the "pure control", compared to the control condition, which is here the omitted category. We run a simple bivariate OLS model where the independent variable is an indicator representing assignment to pure control. We find no differences between the control and pure control conditions.

Table J.1: Difference Between Control and Pure Control Conditions

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Pure Control	−0.172	−0.059	−0.028	0.100	−0.074	0.085	−0.084
	(0.108)	(0.089)	(0.100)	(0.102)	(0.100)	(0.172)	(0.102)
Constant	2.763***	3.277***	2.829***	3.001***	1.656***	2.993***	2.834***
	(0.035)	(0.025)	(0.028)	(0.029)	(0.029)	(0.035)	(0.030)
Observations	1,117	1,351	1,371	1,458	1,398	1,260	1,382
R^2	0.002	0.0003	0.0001	0.001	0.0004	0.0002	0.0005
Adjusted R^2	0.001	−0.0004	−0.001	−0.00002	−0.0003	−0.001	−0.0002
Res. Std. Er.	1.097	0.898	1.008	1.061	1.032	1.205	1.055
F Statistic	2.539	0.437	0.076	0.972	0.538	0.244	0.675
Note:							*p<0.05; **p<0.01; ***p<0.001

K Main Effect With Controls

Table K.1: Main Effect of Any Correction With Controls

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Any Correction	-0.106^{***}	-0.182^{***}	-0.428^{***}	-0.112^{***}	-0.018	-0.413^{***}	-0.116^{***}
	(0.037)	(0.031)	(0.033)	(0.033)	(0.032)	(0.040)	(0.032)
Dissonant Media	0.084^{*}	-0.011	-0.107^{**}	0.059	-0.087^{*}	0.016	0.030
	(0.050)	(0.044)	(0.046)	(0.049)	(0.046)	(0.059)	(0.047)
Congruent Media	0.041	0.101^{**}	-0.119^{***}	0.048	-0.119^{**}	-0.020	0.037
	(0.049)	(0.045)	(0.045)	(0.048)	(0.046)	(0.059)	(0.045)
Copartisan Politician	-0.022	-0.008		0.091^{*}	0.434^{***}	0.091	
	(0.051)	(0.046)		(0.050)	(0.061)	(0.060)	
Outpartisan Politician	-0.190^{***}	-0.271^{***}		-0.297^{***}	-0.141^{***}	-0.157^{**}	
	(0.064)	(0.057)		(0.065)	(0.047)	(0.080)	
Constant	2.747^{***}	3.275^{***}	2.840^{***}	3.008^{***}	1.662^{***}	2.999^{***}	2.785^{***}
	(0.033)	(0.026)	(0.028)	(0.028)	(0.027)	(0.034)	(0.025)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.004	0.013	0.037	0.008	0.015	0.022	0.003
Adjusted R^2	0.003	0.012	0.037	0.007	0.014	0.021	0.002
Res. Std. Er.	1.094	0.943	1.038	1.057	1.007	1.250	1.048
F Statistic	3.599^{***}	13.555^{***}	65.656^{***}	8.161^{***}	15.753^{***}	23.216^{***}	4.443^{***}

Note: *p<0.1; **p<0.05; ***p<0.01

L Comparing the Effects of Different Types of Social Corrections

Table L.1: The Effect of Different Types of Social Corrections

	Dependent variable:						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unsourced Correction	−0.076	−0.126***	−0.348***	−0.101*	−0.013	−0.292***	−0.091*
	(0.045)	(0.037)	(0.041)	(0.043)	(0.039)	(0.049)	(0.039)
Domain Expert Correction	−0.119**	−0.234***	−0.469***	−0.150***	−0.016	−0.483***	−0.125**
	(0.044)	(0.037)	(0.041)	(0.040)	(0.041)	(0.049)	(0.043)
Any Factchecker Correction	−0.117**	−0.209***	−0.521***	−0.065	−0.042	−0.465***	−0.117**
	(0.044)	(0.037)	(0.040)	(0.040)	(0.039)	(0.050)	(0.039)
Constant	2.746***	3.272***	2.827***	3.010***	1.650***	2.999***	2.788***
	(0.033)	(0.026)	(0.028)	(0.027)	(0.027)	(0.034)	(0.025)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.002	0.009	0.039	0.003	0.0002	0.024	0.003
Adjusted R^2	0.001	0.009	0.038	0.002	−0.0004	0.024	0.002
Residual Std. Error	1.095	0.945	1.037	1.060	1.014	1.249	1.048
F Statistic	3.045*	16.280***	68.117***	4.889**	0.398	42.110***	4.338**
Note:							*p<0.05; **p<0.01; ***p<0.001

Table L.2: Comparing Sourced Corrections and Control to Unsourced Corrections (Omitted Category)

	Dependent variable:						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Control	0.100*	0.125***	0.322***	0.068	0.023	0.251***	0.147***
	(0.045)	(0.037)	(0.041)	(0.042)	(0.039)	(0.049)	(0.039)
Expert Correction	−0.036	−0.113**	−0.148***	−0.072	0.001	−0.226***	−0.024
	(0.042)	(0.037)	(0.041)	(0.043)	(0.041)	(0.048)	(0.043)
Any Factchecker Correction	−0.034	−0.088*	−0.201***	0.012	−0.024	−0.208***	−0.016
	(0.042)	(0.037)	(0.041)	(0.043)	(0.039)	(0.049)	(0.039)
Constant	2.663***	3.152***	2.507***	2.932***	1.633***	2.742***	2.687***
	(0.029)	(0.026)	(0.029)	(0.031)	(0.027)	(0.033)	(0.025)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.002	0.009	0.037	0.002	0.0003	0.022	0.004
Adjusted R^2	0.002	0.009	0.036	0.002	−0.0003	0.022	0.004
Residual Std. Error	1.095	0.945	1.038	1.060	1.014	1.250	1.047
F Statistic	3.715*	16.214***	64.747***	3.891**	0.480	38.838***	7.292***
Note:							*p<0.05; **p<0.01; ***p<0.001

Figure L.1: The Effect of Sourced Corrections and Control Condition, compared to Unsourced Correction (Omitted Category)

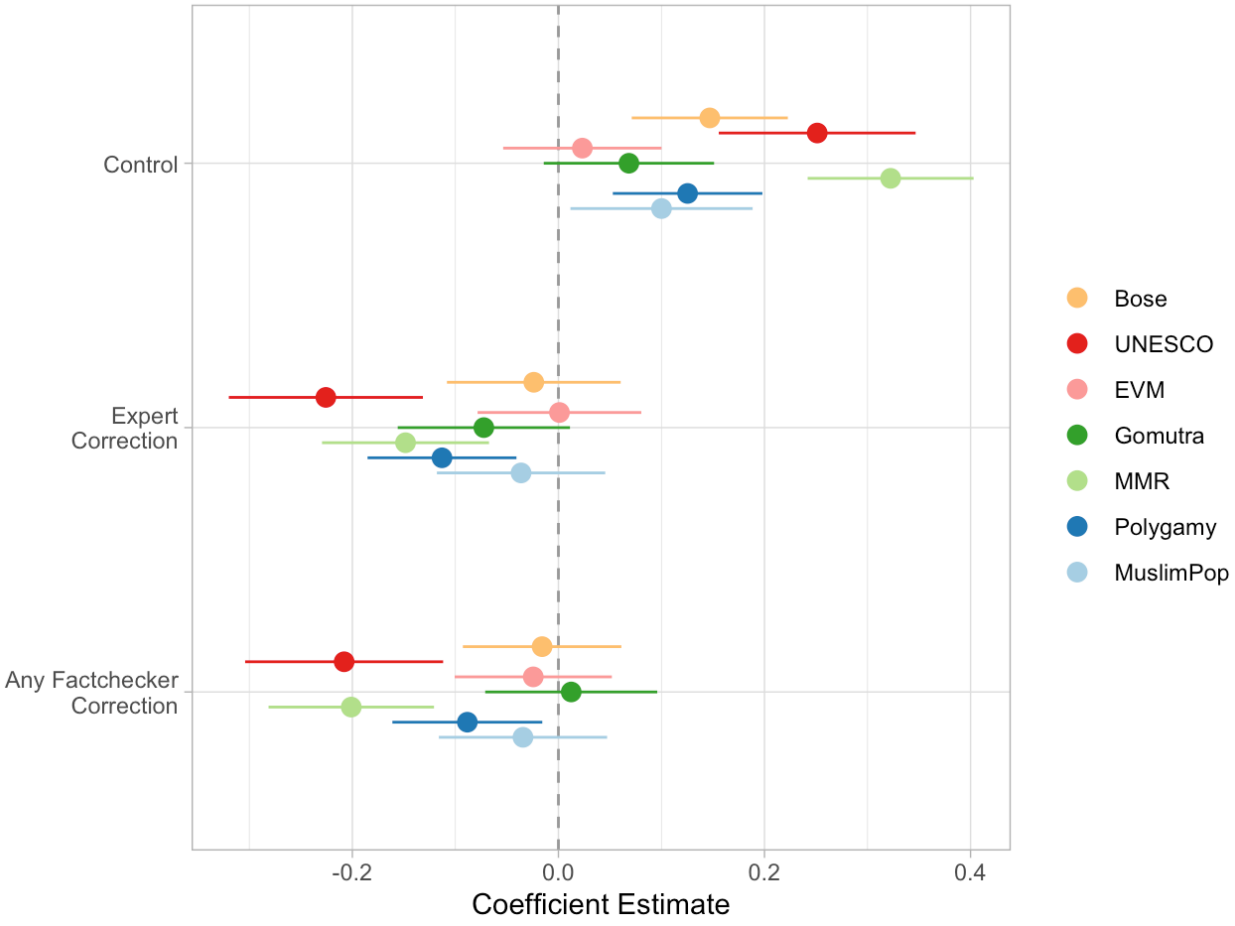


Figure L.2: Comparing Sourced Corrections (pooled) and Control to Unsourced Corrections (Omitted Category)

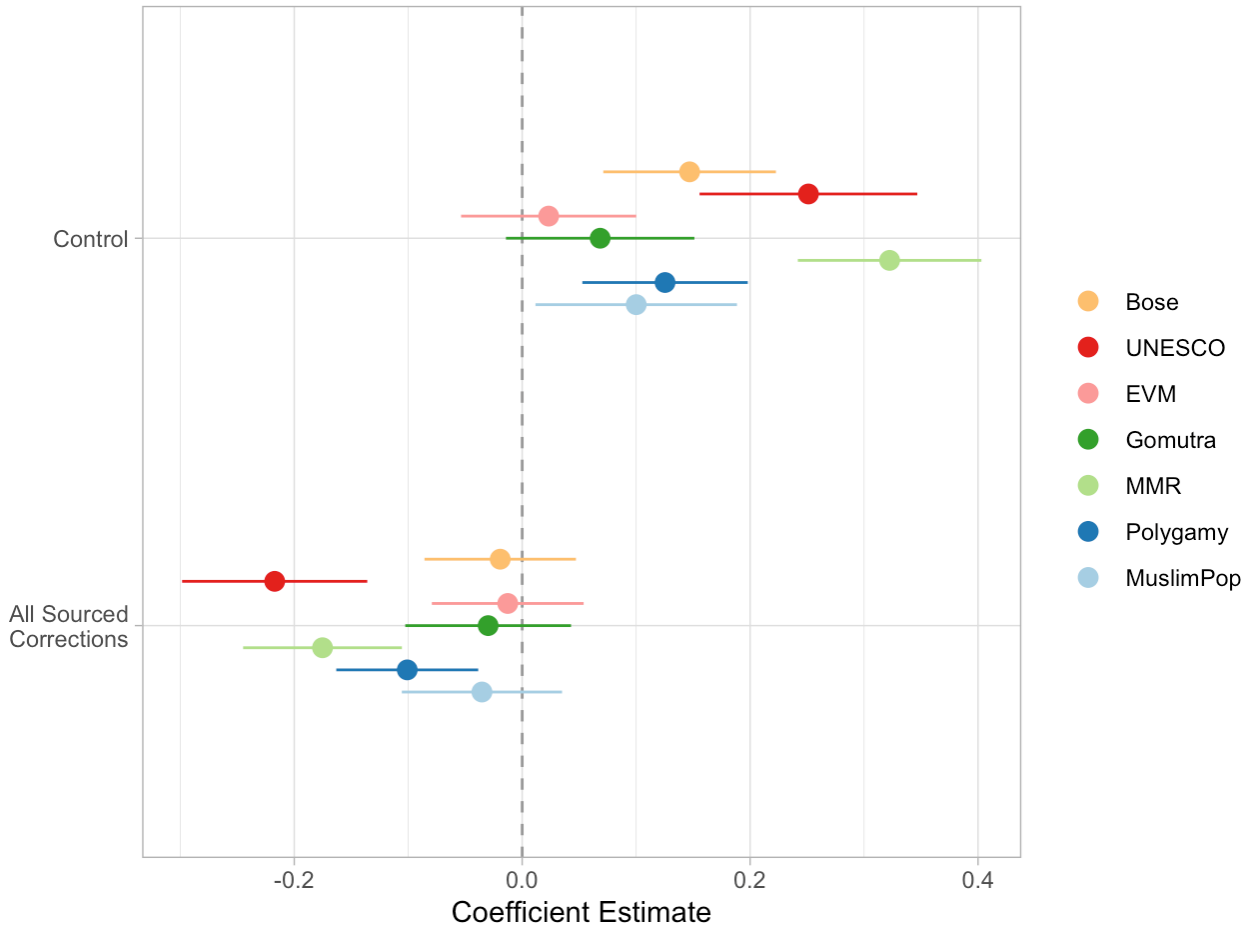


Table L.3: Disentangling by Identity of the Fact-checking Authority

	Dependent variable: Belief in Rumor						
	MuslimPop	Polygamy	MMR	Gomutra	EVM	UNESCO	Bose
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unsourced Correction	−0.076*	−0.126***	−0.348***	−0.101**	−0.013	−0.292***	−0.091**
	(0.045)	(0.037)	(0.041)	(0.042)	(0.039)	(0.049)	(0.039)
Domain Expert	−0.119***	−0.234***	−0.469***	−0.149***	−0.016	−0.483***	−0.125***
	(0.044)	(0.037)	(0.041)	(0.040)	(0.041)	(0.049)	(0.043)
AltNews	−0.079	−0.199***	−0.591***	−0.092	−0.040	−0.487***	−0.171**
	(0.075)	(0.067)	(0.069)	(0.069)	(0.068)	(0.086)	(0.072)
Vishwas News	−0.134*	−0.276***	−0.468***	−0.101	−0.038	−0.367***	−0.075
	(0.074)	(0.064)	(0.069)	(0.071)	(0.067)	(0.090)	(0.074)
Times of India	−0.115	−0.248***	−0.522***	−0.046	−0.051	−0.398***	−0.207***
	(0.077)	(0.065)	(0.071)	(0.073)	(0.068)	(0.085)	(0.072)
Facebook	−0.186**	−0.160**	−0.529***	−0.115	−0.040	−0.547***	−0.102
	(0.074)	(0.066)	(0.073)	(0.070)	(0.066)	(0.088)	(0.072)
WhatsApp	−0.070	−0.158**	−0.494***	0.034	−0.040	−0.529***	−0.028
	(0.073)	(0.065)	(0.072)	(0.074)	(0.070)	(0.091)	(0.071)
Constant	2.746***	3.272***	2.827***	3.010***	1.650***	2.999***	2.788***
	(0.033)	(0.026)	(0.028)	(0.027)	(0.027)	(0.034)	(0.025)
Observations	5,104	5,103	5,061	5,099	5,136	5,109	5,117
R^2	0.002	0.010	0.039	0.004	0.0002	0.025	0.003
Adjusted R^2	0.001	0.009	0.038	0.002	−0.001	0.024	0.002
Res. Std. Er.	1.095	0.945	1.037	1.060	1.014	1.249	1.048
F Statistic	1.589	7.424***	29.487***	2.585**	0.174	18.591***	2.520**

Note: *p<0.1; **p<0.05; ***p<0.01

M Sample Characteristics

Table M.1: Summary Statistics of Key Variables

Statistic	N	Mean	St. Dev.	Min	Median	Max
Age	4,948	29.68	9.43	18	27	76
Male	5,136	0.86	0.34	0	1	1
Education	5,136	6.93	0.97	1	7	8
Hindu	5,136	0.87	0.33	0	1	1
Upper Caste (General)	5,136	0.57	0.49	0	1	1
SC / ST	5,136	0.13	0.34	0	0	1
BJP Partisan	5,136	0.65	0.47	0	1	1
Facebook Use Frequency	5,136	5.40	0.99	1	6	6
WhatsApp Use Frequency	5,136	5.65	0.83	1	6	6

N Benjamini-Hochberg Adjustment

In this table, we present results from a Benjamini-Hochberg adjustment, based on a False Discovery Rate of 5% (0.05).

As seen in the table, the six effects that were significant in Table 2 (main paper) remain significant under Benjamini-Hochberg's adjustment strategy (Benjamini and Hochberg 1995).

Rumor	**Corrective Effect**	**St. Err.**	**p-value**	**Adjusted α**	**Benj.-Hoch Significance**
MuslimPop	0.104	0.037	0.005	0.035	Yes
Polygamy	0.190	0.030	2.91e-10	0.021	Yes
MMR	0.448	0.033	<2e-16	0.014	Yes
Gomutra	0.190	0.033	0.001	0.042	Yes
EVM	0.024	0.032	0.451	0.05	No
Unesco	0.411	0.040	<2e-16	0.014	Yes
Bose	0.109	0.031	0.0004	0.028	Yes