Online Appendices for
Educative Interventions to Combat Misinformation: Evidence From a Field Experiment in India

Contents

A Summary Statistics

Table A.1 provides summary statistics for key variables in this study. Literacy Intervention is a dummy variable indicating random assignment to both treatment groups relative to control. BJP Supporter is a dummy variable indicating respondents' self-reported support for the BJP relative to all other parties. Accurate Priors measures prior beliefs in veracity of news with a battery of four stories (two true and two false); for each story respondents are asked to discern the veracity on a 3-point scale. The variable Accurate Priors calculates the mean accuracy rating across all four stories. Digital Literacy is measured through eight five-point (self-reported) ratings of degree of understanding of WhatsApp-related items. The variable Digital Literacy calculates the mean level of literacy across the eight items. Political Knowledge is measured by a battery of 6 questions of varying difficulty on local and national politics in India; the variable Political Knowledge counts the number of correct answers. WhatsApp Use Frequency measures how frequently respondents use WhatsApp on a 7-point scale ranging from a few times a month to a few times a day. Trust in WhatsApp measures respondents' level of trust in WhatsApp as an accurate medium of receiving news about politics, on a four-point scale.

Table A.1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Median	Max
Literacy Intervention	1,224	0.668	0.471	0	1	1
BJP Supporter	1,224	0.684	0.465	0	1	1
Accurate Priors	1,158	0.695	0.196	0	0.750	1
Digital Literacy	1,224	0.758	0.194	0.083	0.833	1
Political Knowledge	1,224	5.000	1.135	0	5	6
WhatsApp Use Frequency	1,224	6.068	0.952	1	6	7
Trust in WhatsApp	1,224	2.729	0.821	1	3	4
Education	1,224	9.388	2.652	1	9	13
Age	1,224	26.646	9.182	18	24	68
Male	1,224	0.911	0.285	0	1	1
Hindu	1,224	0.837	0.369	0	1	1

B Survey and Sampling Design

The primary sampling unit, the city of Gaya in Bihar, consists of several electoral polling booths (smallest administrative units). Out of the total number of polling booths, a random sample of 85 polling booths were selected (through a random number generator in the statistical framework R) to serve as enumeration areas.

Within each enumeration area, enumerators were instructed to survey 10-12 households following a random walk procedure. This methodology has the benefits of fast implementation and unpredictability of movement and was chosen over traditional listing methods so that enumerators could spend as little time in the field as possible given the potential for electoral violence. It was also chosen over traditional listing methods due to lack of accurate census data and reliable addresses in the area.

Surveying households within each chosen polling booth area involved choosing a starting point and then proceeding along a path, selecting every k^{th} household. I followed the method similar to that used by the Afrobarometer surveys of picking a sample starting point and then choosing a landmark as near as possible to the sample starting point. Landmarks could be street corners, schools, or water sources, and field enumerators were instructed to randomly rotate the choice of such landmarks. From the landmark starting point, the field enumerator walked in a designated direction away from the landmark and selected the tenth household for the survey, counting houses on both the left and the right. Once they left their first interview they continued in the same direction, selecting the next household after another interval of 10. If there were no more houses, the field enumerator turned to the right and kept walking, continuing to count until finding the tenth house. Each field enumerator was assigned to only one polling booth, and hence the paths taken during each selection crossed each household only once.

Once a household is selected, a randomly chosen adult member of the household was chosen to answer our survey questions after they qualified based on pre-conditions.

The three pre-conditions of the survey were (1) access to a personal smartphone (i.e. not a shared household cellphone), (2) connectivity of the phone to working Internet for the past 6 months, (3) usage of WhatsApp on the phone.

Importantly, these qualification conditions resulted in only 20% of all houses knocked on having a respondent who was eligible for the study. This is not atypical for Bihar, where only 20-30% of citizens have access to the internet. Despite this, the study also had a high response rate. Of all those who were eligible for the study, 94.5 percent agreed to participate. The high participation response rate corresponds to research in face-to-face surveys and in developing countries where response rates tend to be typically higher than in developed countries.

Of the 5.5% who refused, enumerator notes suggest that these respondents tended to be older women who (despite having a phone and internet) indicated they would be comfortable if the survey was conducted with a younger member of the household; in some cases they suggested enumerators wait inside the house until a younger member came back home. Once respondents consented to the survey and invited enumerators in their house, no respondent terminated the intervention early or asked that enumerators leave and come back at a different time. Thus, all respondents in the first wave who met the criteria and agreed to the survey completed the intervention in one setting.

The survey pre-conditions ensured that access to WhatsApp and other social media accounts was by the respondent alone, and these restrictions were put into place to ensure that respondents in the study were likely to be exposed to political misinformation over WhatsApp in the months leading up to the election. Sharing mobile phones is especially common among adults in semi-urban and rural India. Further, it is also more common for women than it is for men. Pew survey data from 2019 finds that women are less likely than men to own their own mobile phones, and consequently, significantly more women (20%) than men (5%) report sharing a device with someone else.

These sampling conditions resulted in an uneven age distribution for the study, with

a large proportion of younger respondents under 25. It also resulted in an uneven gender distribution. Focus group discussions with men and women above the age of 45 showed that people in this age group largely did not own their own cellphones; they reported having shared cellphones used by the entire house or not having access to a phone with working Internet at all. Women, particularly, reported using their husbands' cellphones to communicate and did not report owning their own social media accounts. As a result the sample skews younger and male.

C Flyers

Respondents were given flyers as part of the intervention. For treatment group respondents, the front side of the flyer included four false political stories that went viral on social media in the months before the 2019 election. The flyer included the photos / screen grabs associated with these false stories along with an explanation for what the correct version of the story is. The back of the flyer contained 6 general tips to spot misinformation. Enumerators explained each bit of information in the flyer and then finally handed the flyers over to respondents. Treatment 1 flyer has pro-BJP false stories, Treatment 2 flyer has anti-BJP false stories, the control flyer is a placebo and has information on plastic pollution. All materials were in Hindi and the survey and intervention were also administered in Hindi. Below I include English translations of the survey materials.

Figure C.1: Treatment 1 – Pro-BJP Flyer (front and back)

Figure C.2: Treatment 2 – Anti-BJP Flyer (front and back)

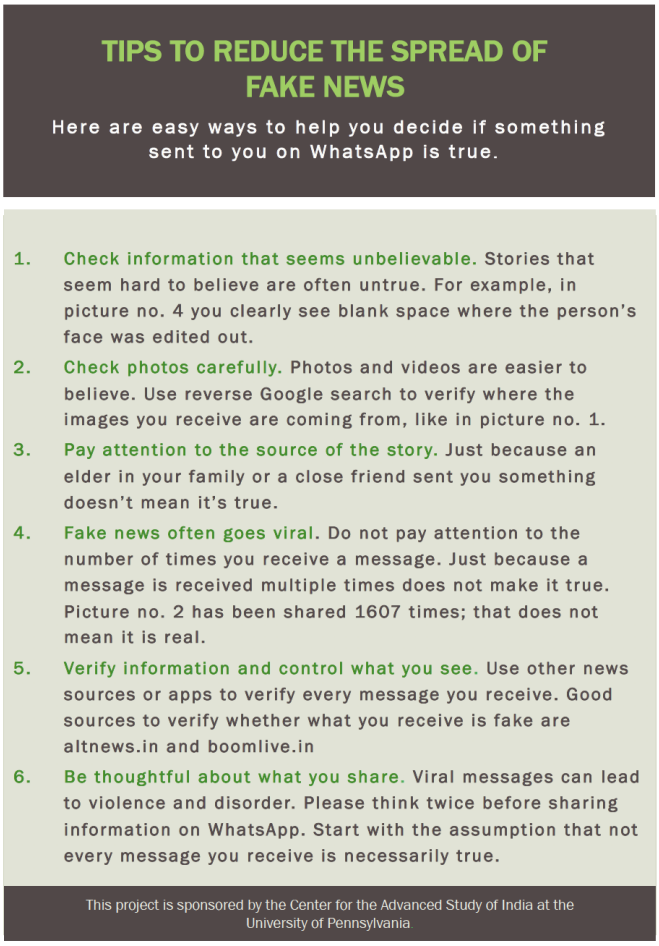


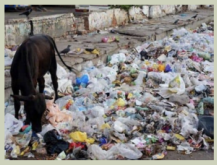
Figure C.3: Placebo Control Flyer (front and back)

Together we can fight plastic usage!

Plastic is harmful for the environment. Here are some ways in which plastic causes damage.

Plastic never goes away. Plastic waste — whether in a river, an ocean, or on land — can persist in the environment for centuries. Most plastic items never fully disappear; they just get smaller and smaller.

Plastic bags cause a serious danger to birds and marine animals that mistake them for food. More than a million birds die each year from plastic pollution. Plastic particles are also swallowed by farm animals or fish.



Plastic affects human health. Because animals mistake it for food, plastic can find their way onto our dinner plates. Exposure to plastic is linked to a number of diseases and ailments.

By clogging sewers and providing breeding grounds for mosquitoes and pests, plastic waste — especially plastic bags — can increase the transmission of vector-borne diseases like malaria.

TIPS TO REDUCE PLASTIC USAGE

Here are easy tips that you can implement to reduce plastic pollution.

1. **Don't use disposable packaging.** Avoid using disposable or single-use plastic. Examples of this are Bisleri water bottles, plastic straws, polythene bags.
2. **Always bring a bag to the shop.** Examples of alternatives to plastic are cloth, jute, and paper.
3. **Reuse.** If you do happen to be in possession of a plastic bag, make sure it is reused for different purposes. Before throwing plastic items, it is important to consider how they can be reused.
4. **Replace plasticware.** Plastic boxes, forks, spoons, plates can be replaced with steel and other materials.
5. **Do not litter.** If you want to throw away plastic, ensure that it is thrown in a bin or trash can. Do not discard plastic into water sources, rivers, or simply out in the open.
6. **Support bans.** Many municipalities in India have enacted bans on single use plastic bags, takeout containers, and bottles. You can support the adoption of such policies in your community.
7. **Spread the word.** Stop others from littering. Stay informed on issues related to plastic pollution and help make others aware of the problem.

This project is sponsored by the Center for the Advanced Study of India at the University of Pennsylvania

D Dependent Variables

To measure key outcomes of interest, respondents were shown a series of fourteen news stories. These stories varied in content, salience, and critically, partisan slant. Half of the stories were pro-BJP in nature and the other half anti-BJP. Each respondent saw all the fourteen stories, but the order in which they were shown was randomized. Table D.1 lists the fourteen stories shown to respondents. Following each story, two primary dependent variables were measured:

1. Perceived accuracy of news stories, with the question "Do you believe this news story is false?" (binary response, 1 if yes, 0 otherwise)

2. Confidence in identification of the story as false or real, with the question "How confident are you that the story is real / false?" (4-point scale, 1 = very confident, 4 = not confident at all)

Table D.1: Dependent Variable Stories

	Story	**Party Slant**	**Veracity**
1	Cow urine cures cancer	Pro-BJP	False
2	Photos of militant bloodshed in Kashmir w/ pro-army message	Pro-BJP	False
3	India has not experienced a single foreign terror attack since 2014	Pro-BJP	False
4	Photoshopped image of war hero in BJP attire	Pro-BJP	False
5	Images of the Indian flag projected onto the Statue of Liberty	Pro-BJP	False
6	Rumor that new Indian notes have tracking chips embedded	Pro-BJP	False
7	Rumor that the govt. has installed CCTV cameras in voting booths	Anti-BJP	False
8	Photoshopped images of BJP workers littering the Ganga river	Anti-BJP	False
9	Rumor that BJP workers use duplicate votes to rig elections	Anti-BJP	False
10	Rumors on lack of policing by govt. leading to child kidnapping	Anti-BJP	False
11	Photoshopped image of govt. built Patel statue developing cracks	Anti-BJP	False
12	Rumors of BJP voters hacking voting machines to rig elections	Anti-BJP	False
13	PM Modi has a new radio show on air called Mann Ki Baat	Pro-BJP	True
14	A recent attack killed 40 Indian CRPF soldiers in Kashmir's Pulwama	Anti-BJP	True

After the fourteen political stories, two additional dependent variables were measured: self-perceived efficacy of the treatment, and self-reported media literacy. Self-perceived efficacy was measured by asking respondents "How confident are you that you can spot false news from real news?" (4-point scale, 1 = very confident, 4 = not confident at all). Media literacy was measured in two ways: trust in news received over WhatsApp (4-point scale); and how frequently they forwarded political messages over WhatsApp (6-point scale). Self-reported literacy and efficacy were measured to determine whether the intervention was successful at generating awareness of the problem of misinformation, arguably demonstrated by decreased trust in WhatsApp and forwarding of political stories. Finally, voter turnout was measured. This was done by asking respondents to show enumerators the index finger of their left hand, which, if they voted, would be marked with purple indelible ink. Because respondents were surveyed within a few days of having voted, the presence of an inked finger is a clean and near-perfect measure of voter turnout. Though this may not be true for instances where respondents refuse to show their ink, in this study every respondent willingly showed enumerators their index finger and no one refused.

Table D.2 is the average treatment effect on the four dependent variables described above. Table D.3 is the heterogeneous effect of party identity on the four dependent variables described above.

Table D.2: Average Treatment Effect on Non-Identification DVs

	Dependent variable:			
	Confidence	Message Checking	Turnout	WhatsApp Trust
	(1)	(2)	(3)	(4)
Literacy Intervention	0.001	−0.015	−0.013	−0.041
	(0.023)	(0.026)	(0.030)	(0.040)
Constant	0.170***	0.246***	0.478***	2.539***
	(0.019)	(0.021)	(0.025)	(0.033)
Observations	1,224	1,224	1,224	1,224
R^2	0.00000	0.0003	0.0002	0.001
Adjusted R^2	−0.001	−0.001	−0.001	0.00004
Residual Std. Error (df = 1222)	0.377	0.425	0.499	0.663
F Statistic (df = 1; 1222)	0.003	0.350	0.192	1.051
Note:				*p<0.1; **p<0.05; ***p<0.01

Table D.3: Heterogeneous Effect of Party on Non-Identification DVs

	Dependent variable:			
	Confidence	Message Checking	Turnout	WhatsApp Trust
	(1)	(2)	(3)	(4)
Literacy Intervention	−0.025	−0.016	−0.038	0.009
	(0.041)	(0.046)	(0.054)	(0.071)
BJP Supporter	0.012	−0.022	0.035	0.103
	(0.040)	(0.045)	(0.053)	(0.070)
Literacy Intervention x	0.039	0.002	0.035	−0.075
BJP Supporter	(0.049)	(0.055)	(0.065)	(0.086)
Constant	0.162***	0.262***	0.454***	2.469***
	(0.033)	(0.037)	(0.044)	(0.058)
Observations	1,224	1,224	1,224	1,224
R^2	0.003	0.001	0.003	0.003
Adjusted R^2	0.0003	−0.002	0.001	0.0004
Residual Std. Error (df = 1220)	0.376	0.425	0.499	0.663
F Statistic (df = 3; 1220)	1.111	0.335	1.377	1.175
Note:				*p<0.1; **p<0.05; ***p<0.01

E Enumerator Fixed Effects

The endline survey to measure the dependent variable was conducted offline (as a paper survey) for field safety reasons. The main dependent variable consisted of 14 stories, but because the survey was conducted offline, the order of appearance of these stories was pre-determined and limited to 3 random orders. A single enumerator only had access to one of the three random orders. Hence as a robustness check, I replicate the main results with enumerator fixed effects.

Table E.1 replicates results for the main effect of the intervention on the outcome. Results are robust to enumerator fixed effects.

Table E.1: Effect of Treatment with Enumerator Fixed Effects

	Dependent variable: Number of Stories Accurately Classified	
	(1)	(2)
Literacy Intervention	−0.013	
	(0.084)	
Literacy + Pro-BJP Fact-Check		−0.005
		(0.097)
Literacy + Anti-BJP Fact-Check		−0.020
		(0.097)
Constant	12.144***	12.144***
	(0.087)	(0.087)
Observations	1,224	1,224
R^2	0.262	0.262
Adjusted R^2	0.260	0.260
Residual Std. Error	1.379 (df = 1220)	1.379 (df = 1219)
F Statistic	144.451*** (df = 3; 1220)	108.258*** (df = 4; 1219)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table E.2 replicates results with enumerator fixed effects for the heterogeneous effect of party identity on a discernment measure of classifying true and false headlines. Results are robust to enumerator fixed effects. Table E.3 splits the key dependent variable into pro-BJP and anti-BJP stories.

Table E.2: Effect of Treatment x Party on Discernment Measure with Enumerator Fixed Effects

	Dependent variable: Number of Stories Accurately Classified
Literacy Intervention	0.364^{**}
	(0.148)
BJP Supporter	0.555^{***}
	(0.146)
Literacy Intervention x BJP Supporter	-0.554^{***}
	(0.179)
Constant	11.768^{***}
	(0.131)
Observations	1,224
R^2	0.271
Adjusted R^2	0.268
Residual Std. Error	1.372 (df = 1218)
F Statistic	90.442^{***} (df = 5; 1218)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table E.3: Effect of Treatment x Party with Enumerator Fixed Effects

	Dependent variable: Number of Stories Identified as False	
	Pro-BJP Stories	Anti-BJP Stories
	(1)	(2)
Literacy Intervention	0.254**	0.077
	(0.103)	(0.093)
BJP Supporter	0.265***	0.327***
	(0.102)	(0.092)
Literacy Intervention x BJP Supporter	−0.384***	−0.120
	(0.125)	(0.112)
Constant	4.608***	5.521***
	(0.092)	(0.082)
Observations	1,224	1,224
R^2	0.258	0.139
Adjusted R^2	0.255	0.135
Residual Std. Error (df = 1218)	0.958	0.860
F Statistic (df = 5; 1218)	84.543***	39.252***
Note:		*p<0.1; **p<0.05; ***p<0.01

F Correlates of Misinformation

Independent of the literacy intervention, it is descriptively interesting for the understudied context of India to understand who is more likely to consume misinformation and more likely to be able to identify news as false. I consider the main effect of several demographic and pre-treatment variables on ability to identify misinformation. The results are presented in Table F.1. For all dependent variable stories taken together, BJP partisans are better at identifying false stories as compared to their non-BJP partisan counterparts. Further, as expected, accurate prior beliefs are more likely to aid in identifying misinformation. However, higher levels of digital literacy were associated with worse levels of discernment, underscoring that greater knowledge of how to use WhatsApp may lead to more vulnerability to misinformation in this context. However, those who report using WhatsApp more often are more likely to be able to identify misinformation. Interestingly, higher levels of trust in WhatsApp do not correlate with identification of false stories, suggesting that familiarity with the medium itself can make people more vulnerable to misinformation and consequently more likely to share false stories.

With respect to demographic variables, increase in age and education is associated with a higher capacity to identify misinformation.

Table F.1: Correlates of Misinformation

	Dependent variable: Number of Stories Accurately Classified
Literacy Intervention	-0.072
	(0.097)
BJP Supporter	0.242^{**}
	(0.115)
Accurate Priors	0.682^{***}
(Higher = more accurate)	(0.236)
Digital Literacy	-1.332^{***}
(Higher = more literate)	(0.258)
Political Knowledge	-0.069
(Higher = more knowledge)	(0.047)
WhatsApp Use Frequency	0.180^{***}
(Higher = more usage)	(0.049)
Trust in WhatsApp	-0.015
(Higher = more trust)	(0.059)
Education	0.045^{**}
	(0.019)
Age	0.019^{***}
	(0.005)
Hindu	-0.342^{**}
	(0.148)
Male	0.295^{*}
	(0.167)
Constant	10.442^{***}
	(0.447)
Observations	1,158
R^2	0.071
Adjusted R^2	0.062
Residual Std. Error	1.542 (df = 1146)
F Statistic	7.979^{***} (df = 11; 1146)
Note:	*p<0.1; **p<0.05; ***p<0.01

G Age and Digital Literacy

I explore further the relationship between age, misinformation, and digital literacy. The tables below look at age as variable. In Table G.1, I demonstrate that older respondents are better at discernment. However in Table G.2, I find that older respondents have lower levels of digital literacy, demonstrating that despite having better digital literacy skills, younger respondents are worse are identifying false stories.

Table G.1: Effect of Age on Identification of Misinformation

	Dependent variable: Number of stories accurately classified
	(1)
Age (Continuous)	0.019***
	(0.005)
BJP Supporter	0.242**
	(0.115)
Digital Literacy	−1.335***
(Higher = more literate)	(0.258)
Accurate Priors	0.675***
(Higher = more accurate)	(0.236)
Political Knowledge	−0.069
(Higher = more knowledge)	(0.047)
WhatsApp Use Frequency	0.181***
(Higher = more usage)	(0.049)
Trust in WhatsApp	−0.014
(Higher = more trust)	(0.059)
Education	0.045**
	(0.019)
Male	0.295*
	(0.167)
Hindu	−0.340**
	(0.148)
Constant	10.400***
	(0.443)
Observations	1,158
R^2	0.071
Adjusted R^2	0.063
Residual Std. Error	1.542 (df = 1147)
F Statistic	8.725*** (df = 10; 1147)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table G.2: Effect of Age on Digital Literacy

	Dependent variable: Digital Literacy (Higher = more literate)
	(1)
Age (Continuous)	-0.002^{***}
	(0.001)
BJP Supporter	0.020
	(0.013)
Accurate Priors	-0.063^{**}
(Higher = more accurate)	(0.027)
Political Knowledge	0.051^{***}
(Higher = more knowledge)	(0.005)
WhatsApp Use Frequency	0.006
(Higher = more usage)	(0.006)
Trust in WhatsApp	0.024^{***}
(Higher = more trust)	(0.007)
Education	0.007^{***}
	(0.002)
Male	0.079^{***}
	(0.019)
Hindu	-0.028
	(0.017)
Constant	0.376^{***}
	(0.049)
Observations	1,158
R^2	0.179
Adjusted R^2	0.173
Residual Std. Error	0.177 (df = 1148)
F Statistic	27.809^{***} (df = 9; 1148)
Note:	*p<0.1; **p<0.05; ***p<0.01

I now consider whether the literacy intervention worked better depending on age or digital literacy. In Table G.3 I interact the treatment with age and digital literacy, and do not find an interaction effect.

Table G.3: Effect of Treatment x Age and Digital Literacy

	Dependent variable: Number of Stories Accurately Classified	
	(1)	(2)
Literacy Intervention	0.386	0.349
	(0.306)	(0.382)
Age (Continuous)	0.032***	
	(0.009)	
Literacy Intervention x Age	−0.015	
	(0.011)	
Digital Literacy (Higher = more literate)		−0.984**
		(0.394)
Literacy Intervention x Digital Literacy		−0.454
		(0.489)
Constant	10.797***	12.378***
	(0.257)	(0.306)
Observations	1,224	1,224
R^2	0.016	0.025
Adjusted R^2	0.014	0.022
Residual Std. Error (df = 1220)	1.592	1.585
F Statistic (df = 3; 1220)	6.673***	10.270***
Note:		*p<0.1; **p<0.05; ***p<0.01

H True Stories

The outcome measure for this study comprised of more false stories than true (rather than a 50-50 split between true and false stories). This was done to maximize reducing belief in as many false stories as possible. However, several steps were taken to ensure that the imbalance of true vs. false stories did not affect the efficacy of the treatment. Before measuring the outcomes, respondents were told that some of the stories were false and some true, likely reducing the urge to default to the stories being false. Further, with the comprehension check, respondents were not only asked whether stories were true or false but were also asked how they identified the veracity of these stories. Importantly, a majority of respondents in the treatment groups said that their responses were motivated by enumerators teaching them about these stories during the household visit, rather than having learnt about the stories on the news or through a friend. Further, enumerators were instructed for this question to not read out response options aloud, but to allow respondents to organically speak about their views on the false stories in a way that minimized the ability of respondents to provide socially desirable answers. Thus, much care was taken in the experiment to ensure that the skew towards false stories would not impact respondents' answers.

I now analyze whether the treatment worked for the two true stories alone. Results are in Table H.1. I find that the perceptions of veracity of these stories did not depend on the treatment. However, respondents accurately classified a high proportion of the true stories, 76% and 95% respectively.

Table H.1: Identification of True Stories

	Dependent variable: Accurate Identification	
	1st True Story	2nd True Story
Literacy Intervention	−0.009	0.005
	(0.026)	(0.013)
Constant	0.776***	0.951***
	(0.021)	(0.010)
Observations	1,224	1,224
R^2	0.0001	0.0001
Adjusted R^2	−0.001	−0.001
Residual Std. Error (df = 1222)	0.421	0.209
F Statistic (df = 1; 1222)	0.134	0.171
Note:		*p<0.1; **p<0.05; ***p<0.01

I BJP Partisanship

I explore in this section the finding that while BJP partisans are more likely to identify rumors correctly, pro-BJP rumors are also more likely to be believed. To better understand this result, I demons rate that BJP supporters' better identification is driven by their ability to identify anti-BJP stories as fake. In Table I.1 below, I list the rate of identification (average number of stories correctly identified) for BJP and non-BJP supporters for the two categories of stories that make up the DV, pro-BJP and anti-BJP rumors. The results from a two-sample t-test demonstrate that for pro-BJP stories, the difference in means is not different from 0. But for anti-BJP stories, the true difference in means is significantly greater than 0, with BJP respondents correctly identifying stories at a higher rate. Thus the finding that BJP respondents identify stories at a higher rate is driven only by their identification of anti-BJP messages as false.

Table I.1: T-test for Discernment by Party ID

	Mean for BJP	**Mean for non-BJP**	**T**	**p-value**
Pro-BJP Stories	4.55	4.59	0.69	0.488
Anti-BJP Stories	5.41	5.19	-3.63	0.000

Parsing this result further, I now examine whether salience of stories is linked with party-congruent beliefs. I examine whether there is a partisan divide in highly believed stories. To do so, I limit the dependent variable to the top two most believed stories (belief close to 50% of the sample); both stories are pro-BJP in slant. I find that for these stories, BJP supporters do significantly worse at discernment, i.e., they are significantly more likely to think that these false stories are true. I report this result in Table I.2.

Table I.2: Discernment of Top Two Believed Stories

	Dependent variable: Number of Stories Accurately Classified
	(1)
BJP Supporter	-0.130^{***}
	(0.045)
Constant	1.171^{***}
	(0.037)
Observations	1,224
R^2	0.007
Adjusted R^2	0.006
Residual Std. Error	0.727 (df = 1222)
F Statistic	8.458^{***} (df = 1; 1222)
Note:	*p<0.1; **p<0.05; ***p<0.01