Appendix:
Misinformation and Support for Vigilantism

Contents

A Sampling Strategy **2**

B Experimental Design **3**

C The Media Landscape in India and Pakistan **7**

D Baseline Belief in Rumors **8**

E Main Effects for Sunni Subsample **9**

F Summary Statistics **11**

G District Fixed Effects **13**

H Standardized Dependent Variables **15**

I Effect of Positionality **17**

J Robustness Checks for Table 4 **20**

K Ethics **24**

L Enumerator Fixed Effects **27**

A Sampling Strategy

Our surveys were conducted in the state of Uttar Pradesh (UP) in India and Punjab province in Pakistan. We chose to focus on these two regions not only because both states/provinces are each countries' most populous, but also because both have seen a number of acts of vigilante violence in recent years.

In UP, we selected three districts of the Western part of the state in which we could identify that specific vigilante incidents had occured in recent years. Available (admittedly imperfect) data suggested that the Western part of the state was a hotspot for vigilante events.[1] Within each district we then randomly selected three assembly constituencies (AC). Within each AC, we randomly selected 10 polling booth areas. In each polling booth area, we randomly selected 20 respondents from voter lists, for a total of 200 in each selected AC, 600 in each district and 1800 overall. In each case, we also randomly drew a replacement list in case initially targeted respondents were unable to participate after repeated tries.

In Punjab, Pakistan, we selected one district in each of south, central, and north Punjab to allow for greater variation within that province. Within each district, we selected two constituencies, with a focus on more urban locales. Within each constituency, the survey team created 100 square meter grids for which they calculated the density and then selected 50 grids within each unit randomly or on the basis of density. From there, we used a random household selection method by dropping 50 random pins and conducting 10 random interviews (using left-hand rule) against each randomly dropped pin.

Pre-tests were conducted prior to piloting the survey, in which enumerators were able to double-check that the audio-visual stimulus was working as intended and ensure the vignette wording was capturing the desired intent. Following these pre-tests, a pilot of 100 respondents in each country was carried out. Enumerators were trained directly by the authors.

[1] See for instance, the maps in https://acleddata.com/2021/05/03/cow-protection-legislation-and-vigilante-violence-in-india/

B Experimental Design

Text of the vignette scenarios

Below is an illustrative example of what the experimental scenarios looked like in both India and Pakistan. For example, in one scenario in India, respondents were shown Figure B1 around the time the enumerators began the audio. The rationale for this visual component was threefold. First, it allowed us to reinforce the realism of the prompt, insofar as a headline lent credibility to the fact that an incident had actually happened. Second, practically, it allowed us to maintain the focus of the respondents on the story being described in the audio. Third, it helped us ensure that respondents did not become too alarmed or agitated – a pressing ethical concern in this context – before they listened to the entire audio, as the visual headline stated that (a) the incident had not happened in their immediate surrounding and (b) that no one had been killed over it; in that sense, it limited potential implications of hearing about the incident.

Figure B1: Example of Social Media Post Seen By Respondents (India)

Note: We deliberately hid the name of the author of the purported post, as well as other details to limit source effects. Text says "Mob angry over viral claim that Muslim man abducted Hindu woman to marry her burns buildings in [town name hidden]."

The audio started with the following:

Let me tell you about the incidents that are described in this social media post. In this area, last January, people heard a rumor on social media. The rumor said that a Muslim man had

abducted a Hindu woman in order to marry her. One day, when the Muslim man returned home from work, an angry mob attacked the neighborhood of the Muslim man. In the ensuing destruction, many buildings were burnt, as you can see in the photo.

Next, information (or no information) about state positionality followed. In the treatment vignettes, this included information that the state either intervened to investigate the alleged crime or intervened to punish the mob.

Positionality treatment 1: After the incident, local police sprung into action. They decided to investigate whether a crime had actually been committed – that is, whether the Muslim man had actually abducted the Hindu woman – as the rumor suggested.

Positionality treatment 2: After the incident, local police sprung into action. They started to round up the vigilante mob and arrested them for unlawfully taking justice into their own hands.

The correction treatment then included the following language:

Subsequent to this incident, reporters from several main Hindi news channels – such as Aaj Tak, Zee News, and NDTV – spent a month thoroughly investigating the story. They interviewed dozens of local stakeholders with knowledge of the case, such as the police, eyewitnesses, neighbors, netas. Finally their investigation found that the initial rumor was actually baseless: it was a love marriage, and the Muslim man had not kidnapped the Hindu woman, as the post claimed. The mob violence took place because of false information.

For the cow vigilante scenario, we added an additional elite rhetoric treatment with the following prompt, followed by a screenshot of the Tweet presented in the main text of the manuscript:

Following this incident, several politicians and leaders reacted to the violence and put out statements. For example, this is what Prime Minister Narendra Modi said on twitter.

Below is an example scenario from Pakistan about posting blasphemous content online.

Let me tell you about the incidents that are described in this social media post. In this area, last January, people heard a rumor on social media. The rumor said that that a man had posted blasphemous content on Facebook. One day, young men gathered in the street of the accused man and attacked his neighborhood. In the ensuing destruction, many buildings were burnt, as you can see in the photo.

Positionality 1: After the incident, local police sprung into action. They decided to investigate whether a crime had indeed been committed—that is, whether blasphemous content had

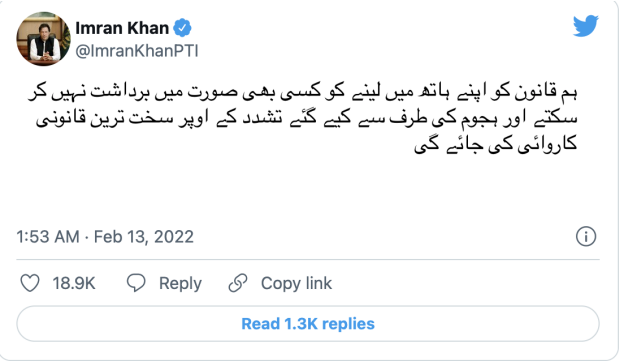
been posted online—as the rumor had suggested.

Positionality 2: After the incident, local police sprung into action. They started to round up the vigilante mob and arrested them for unlawfully taking justice into their own hands.

Correction: Subsequent to the incident, reporters from numerous TV channels, including ARY, GEO, and Samaa TV, spent a month thoroughly investigating the story in the area. They interviewed dozens of local stakeholders with knowledge of the case, such as the police, eyewitnesses, and neighbors. Their investigation found that the initial rumor was actually baseless: no blasphemous content was posted by the man, as social media posts had claimed. The mob violence took place because of false information.

Following this incident, several politicians and leaders reacted to the violence and put out statements. For example, this is what then Prime Minister Imran Khan said on twitter.

Figure B2: Example of Elite Message (Pakistan)



Text reads: "We have zero tolerance for anyone taking the law into their own hands & mob lynchings will be dealt with full severity of the law."

Order of attributes

Within each scenario, we pre-decided the order in which the different attributes—rumor corrections, state positionality, and elite rhetoric—were delivered. Namely, in vignettes with both the correction treatment and the positionality treatments, the correction treatment always appeared second. When it was combined with the elite rhetoric treatment, the correction treatment appeared first. In vignettes with all three treatments, the order was consistent: state positionality followed by corrections followed by elite rhetoric. We proceeded in this way as we believed this to be closest to the logical chronology of what would happen in response to an actual event. The state would immediately have to react (or not). By contrast, journalistic investigations would take more time. We believed this to be a necessary step to ensure the chronological coherence of the stories.

Comprehension check

We included a comprehension check question immediately after each audio recording ended and before outcomes were measured. This single item measured responses to a simple factual question about the vignette, designed to ensure both that respondents were comprehending the audio and paying attention. For example, in the rumor only, no correction condition, respondents were asked to identify the type of crime which led to the mob attacking the individual (with options including the communal or religious crime in question, and wrong answers including tax fraud, bank robberies, etc). If respondents failed this check, our survey protocol instructed enumerators to summarize the the main takeaways of the audio verbally to respondents. 96% passed the check in Pakistan while 89% passed it in India; these differences might owe to vastly different levels of education in both samples.

Item wording for outgroup attitudes

In order to test crucial heterogeneous effects by degree of pre-existing (in)tolerance and (dis)trust of minorities, we measured pre-treatment attitudes towards outgroups through the following survey items (words in parentheses reflect terms used in the Pakistan component of the survey).

- I am going to name a number of organizations and groups. For each one, could you tell me how much trust you have in them: is it a great deal of trust, quite a lot of trust, not very much trust or none at all? Your neighbors, the courts in India (Pakistan), the National government, Hindus (Muslims), Muslims (non-Muslims), the media
- How comfortable are you having close friends that are Muslims (non-Muslim)? Very comfortable, Somewhat comfortable, Not very comfortable, Not at all comfortable

C The Media Landscape in India and Pakistan

Digital media has been rapidly gaining ground in India. A study by the Reuters Institute Digital News Report 2021 stated that 58 percent of Indian respondents used social media for news, and 66 percent used online news websites or apps. Additionally, the report noted that the COVID-19 pandemic had accelerated the trend of people consuming news online. However, the dominant medium of news consumption in India undoubtedly remains television. According to a report by the Broadcast Audience Research Council (BARC), more than two-thirds of the Indian population (approximately 836 million people) watched TV every week, with news channels being among the most-watched.

It is widely recognized that some news channels in India have political affiliations and biases towards particular parties or ideologies. Certain news channels in India have been accused of being biased towards the current government and Prime Minister Narendra Modi (Zee News would be a good example), while others are seen as either pro-opposition or more neutral (NDTV at the time of fielding our survey is one example). Some critics have accused certain channels of promoting a pro-government narrative and ignoring or downplaying issues that may be critical of the government. There have also been allegations of some news channels using inflammatory language, sensationalizing news stories, and spreading false or misleading information - including on the topics we manipulate here - to promote a particular political agenda. Despite this, we found several examples where news channels engage in the type of investigative debunking we describe in our main treatment.

In Pakistan, television remains the dominant medium of news consumption. There are dozens of private channels with one state broadcast channel, the Pakistan Television Corporation (PTV). Some television stations are perceived to be partial to the military establishment, while others are thought to be biased towards some political parties; the primary fault-line is between Geo TV and ARY TV. However, blasphemy is not a particularly salient issue which cuts across television networks. As such, it would be perceived as plausible that the main television channels mentioned in the experimental treatment–ARY, Geo, and Samaa TV—would each have conducted investigations to explore rumors leading to violence.

D Baseline Belief in Rumors

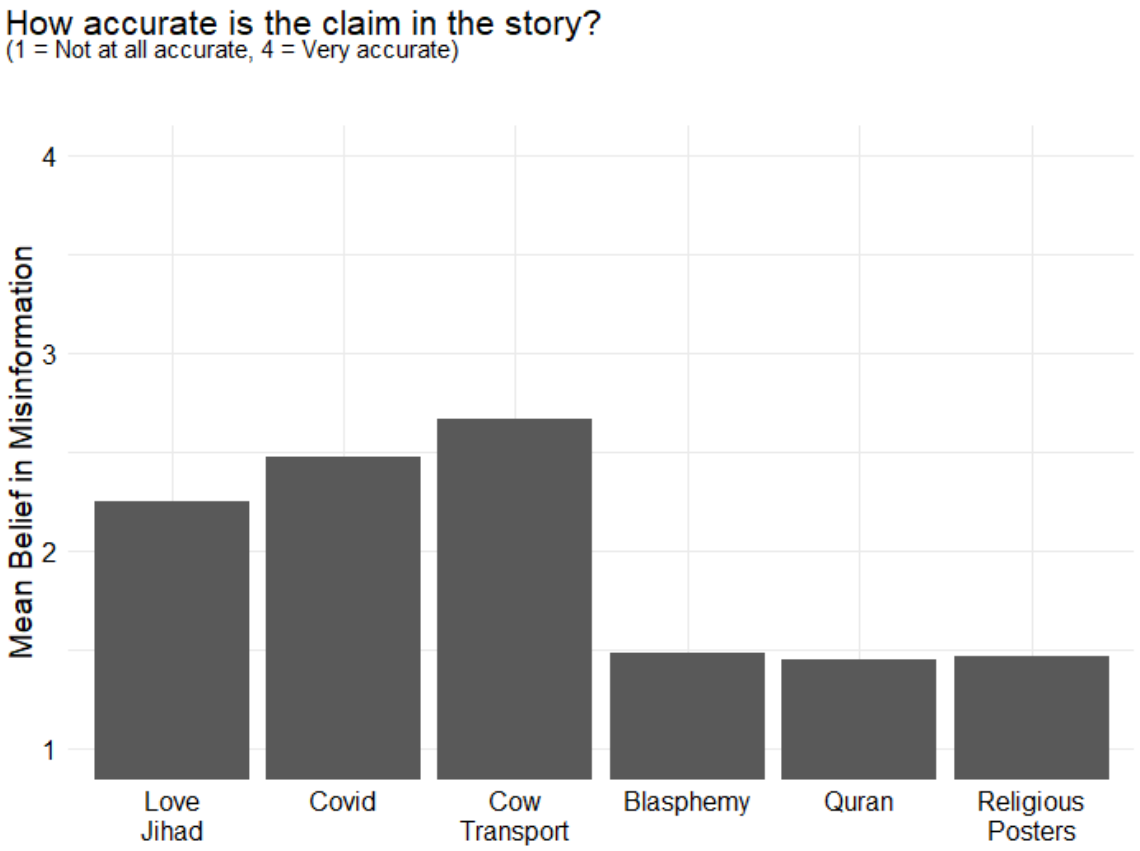
Baseline belief in the underlying rumors which led to the vigilante acts varied significantly between India and Pakistan.

As the graph below shows, respondents in India were overall more likely to believe that the alleged crime had occurred, while those in Pakistan were less likely to do so.

Because the rumors we selected were primarily chosen with an eye to realism - they were, by design, to resemble recent rumors that had led to incidents in the two contexts - they were not meant to be equally salient or common across the two contexts.

Accordingly, there are a number of reasons why we might have seen a difference between the two countries; future research should study which of these is the most likely explanation. First, it is possible that the target outgroup (Muslims in India, non-Muslims in Pakistan) was more clearly identified (and/or subject to rumors), more salient, or provided a more credible target for a rumor in our India sample than in our Pakistan sample. Second, it is possible that a difference in the demographics of the two samples (more urban vs. rural, for example) explains the difference, or a difference in levels of education.

Figure D3: Belief in Misinformation



E Main Effects for Sunni Subsample

Table E1: Main Effects of Treatments on Support for Vigilantism Among Sunnis

	Support for Vigilantism	
	Own Support	Neighborhood Support
	Sunni	*Sunni*
Correction Treatment	-0.432^{***}	-0.377^{***}
	(0.052)	(0.056)
Positionality 1: Police	0.002	-0.057
Investigates Crime	(0.060)	(0.067)
Positionality 2: Police	0.006	-0.067
Prosecutes Vigilantes	(0.062)	(0.066)
Crime: Burning Quran	-0.018	-0.064
	(0.047)	(0.052)
Crime: Tearing Down	-0.025	-0.056
Posters	(0.045)	(0.051)
Constant	2.428^{***}	2.795^{***}
	(0.067)	(0.069)
Observations	4,080	4,004
R^2	0.019	0.013
Adjusted R^2	0.018	0.011
Residual Std. Error	1.543 (df = 4074)	1.705 (df = 3998)
F Statistic	16.044^{***} (df = 5; 4074)	10.215^{***} (df = 5; 3998)

Note: *p<0.05; **p<0.01; ***p<0.001

Table E2: Main Effects of Treatments on Support for Punishing Mob Among Sunnis

	Should the vigilante mob be punished?
	Sunni
Correction Treatment	0.163***
	(0.022)
Positionality 1: Police	0.008
Investigates Crime	(0.025)
Positionality 2: Police	−0.001
Prosecutes Vigilantes	(0.026)
Crime: Burning Quran	0.025
	(0.020)
Crime: Tearing Down	0.032
Posters	(0.020)
Constant	3.467***
	(0.028)
Observations	4,027
R^2	0.016
Adjusted R^2	0.015
Residual Std. Error	0.650 (df = 4021)
F Statistic	12.959*** (df = 5; 4021)
Note:	*p<0.05; **p<0.01; ***p<0.001

F Summary Statistics

Table F3: Summary of India Variables

Statistic	N	Mean	St. Dev.	Min	Max
Belief in Misinformation	5,505	2.461	1.330	1	4
Support for Vigilantism (Self)	4,365	4.681	3.328	1	10
Support for Vigilantism (Neighbors)	4,265	5.146	3.265	1	10
Support for Punishing Mob	5,499	3.290	1.089	1	4
Correction Treatment	5,508	0.496	0.500	0	1
Positionality 1	5,508	0.331	0.471	0	1
Positionality 2	5,508	0.333	0.471	0	1
Elite Messaging	5,508	0.164	0.370	0	1
Feelings Toward Modi	5,508	78.647	30.053	0	100
Trust in Religious Majority	5,472	3.578	0.786	1	4
Trust in Religious Minority	5,211	1.948	1.078	1	4
Trust in Media	5,136	2.897	1.105	1	4
Trust in Republic TV	3,279	3.752	1.360	1	5
Trust in Aaj Tak	3,987	4.135	1.194	1	5
Trust in Dainik Jagran	3,786	4.168	1.168	1	5
Trust in Zee News	3,279	3.762	1.376	1	5
Trust in NDTV	2,736	3.477	1.465	1	5
Trust in Amar Ujala	3,918	4.341	1.011	1	5
Trust in Dainik Bhaskar	2,730	3.684	1.408	1	5
Trust in Hindustan Dainik	3,072	3.921	1.248	1	5
Trust in News 18	2,817	3.742	1.344	1	5
Pre-Treat Support for Punishing Accused	5,487	0.032	0.175	0	1
Support for Free Speech	5,508	0.862	0.345	0	1
Support for Due Process	5,508	0.802	0.398	0	1
Crim. Should Be Severely Punished	5,508	0.937	0.243	0	1
Laws Should Favor Rel. Maj.	5,508	0.439	0.496	0	1
Army Should Overthrow Incomp. Govt.	5,508	0.570	0.495	0	1
Support for Violent Protest	5,508	0.231	0.422	0	1

Table F4: Summary of Pakistan Variables

Statistic	N	Mean	St. Dev.	Min	Max
Belief in Misinformation	4,413	1.467	0.645	1	4
Support for Vigilantism (Self)	4,490	2.211	1.564	1	10
Support for Vigilantism (Neighbors)	4,408	2.525	1.722	1	10
Support for Punishing Mob	4,427	3.571	0.655	1	4
Correction Treatment	4,500	0.506	0.500	0	1
Positionality 1	4,500	0.339	0.473	0	1
Positionality 2	4,500	0.330	0.470	0	1
Elite Messaging	4,500	0.165	0.371	0	1
Feelings Toward Khan	4,500	57.049	41.901	0	100
Trust in Religious Majority	4,500	3.645	0.611	1	4
Trust in Religious Minority	4,491	1.708	0.818	1	4
Trust in Media	4,488	2.375	0.854	1	4
Pre-Treat Support for Punishing Accused	4,500	0.081	0.272	0	1
Support for Free Speech	4,500	0.714	0.452	0	1
Support for Due Process	4,500	0.643	0.479	0	1
Crim. Should Be Severely Punished	4,500	0.422	0.494	0	1
Laws Should Favor Rel. Maj.	4,500	0.553	0.497	0	1
Army Should Overthrow Incomp. Govt.	4,500	0.121	0.327	0	1
Support for Violent Protest	4,500	0.044	0.205	0	1
Support for Due Process	4,500	0.643	0.479	0	1
Sunni	4,500	0.909	0.288	0	1

G District Fixed Effects

Adding fixed effects for districts to the model did not change the results significantly.

Table G5: Main Effects on Support for Vigilantism with District FEs

	Support for Vigilantism			
	Own Support	Neighborhood Support	Own Support	Neighborhood Support
	India		*Pakistan*	
Variables				
Correction Treatment	-0.419***	-0.404***	-0.472***	-0.407***
	(0.102)	(0.104)	(0.050)	(0.054)
Positionality 1: Police	0.065	0.094	-0.007	-0.057
Investigates Crime	(0.123)	(0.122)	(0.058)	(0.063)
Positionality 2: Police	0.008	-0.127	-0.013	-0.071
Prosecutes Vigilantes	(0.123)	(0.123)	(0.059)	(0.063)
Crime: Cow Transport	0.341***	0.401***		
	(0.091)	(0.093)		
Crime: Love Jihad	-0.209**	-0.086		
	(0.091)	(0.092)		
Crime: Burning Quran			0.014	-0.034
			(0.045)	(0.050)
Crime: Tearing Down			-0.031	-0.056
Posters			(0.043)	(0.048)
Fixed-effects				
District	Yes	Yes	Yes	Yes
Fit statistics				
Observations	4,365	4,265	4,490	4,408
R^2	0.01251	0.00937	0.03074	0.01537
Within R^2	0.00865	0.00885	0.02314	0.01444

Clustered (Respondent ID) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table G6: Main Effects on Support for Punishing Mob with District FEs

	Should the vigilante mob be punished?	
	India	*Pakistan*
Variables		
Correction Treatment	0.064**	0.173***
	(0.029)	(0.021)
Positionality 1: Police	0.034	0.010
Investigates Crime	(0.036)	(0.024)
Positionality 2: Police	0.094***	0.014
Prosecutes Vigilantes	(0.036)	(0.024)
Crime: Cow Transport	-0.168***	
	(0.028)	
Crime: Love Jihad	0.005	
	(0.027)	
Crime: Burning Quran		0.015
		(0.020)
Crime: Tearing Down		0.030
Posters		(0.019)
Fixed-effects		
District	Yes	Yes
Fit statistics		
Observations	5,499	4,427
R^2	0.01035	0.03072
Within R^2	0.00763	0.01814

Clustered (Respondent ID) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H Standardized Dependent Variables

To standardize dependent variables, we subtracted the sample mean from each observation and divided it by the standard deviation. Results hold.

Table H7: Main Effects of Treatments on Support for Vigilantism (SD Units)

	Support for vigilantism			
	Own Support *India*	Neighborhood Support *India*	Own Support *Pakistan*	Neighborhood Support *Pakistan*
Correction Treatment	−0.130***	−0.125***	−0.301***	−0.236***
	(0.031)	(0.032)	(0.032)	(0.031)
Positionality 1: Police	0.022	0.030	−0.002	−0.033
Investigates Crime	(0.037)	(0.037)	(0.037)	(0.037)
Positionality 2: Police	0.004	−0.039	−0.004	−0.040
Prosecutes Vigilantes	(0.037)	(0.038)	(0.038)	(0.036)
Crime: Cow Transport	0.103***	0.122***		
	(0.027)	(0.028)		
Crime: Love Jihad	−0.063*	−0.026		
	(0.027)	(0.028)		
Crime: Burning Quran			0.009	−0.020
			(0.029)	(0.029)
Crime: Tearing Down			−0.020	−0.033
Posters			(0.027)	(0.028)
Constant	0.041	0.032	0.158***	0.161***
	(0.038)	(0.038)	(0.041)	(0.038)
Observations	4,365	4,265	4,490	4,408
R^2	0.009	0.009	0.023	0.014
Adjusted R^2	0.008	0.008	0.022	0.013
Residual Std. Error	0.996 (df = 4359)	0.996 (df = 4259)	0.989 (df = 4484)	0.993 (df = 4402)
F Statistic	7.851*** (df = 5; 4359)	7.674*** (df = 5; 4259)	20.888*** (df = 5; 4484)	12.822*** (df = 5; 4402)

Note: *p<0.05; **p<0.01; ***p<0.001

Table H8: Main Effects of Treatments on Support for Punishing Mob (SD Units)

	Should the vigilante mob be punished?	
	India	*Pakistan*
Correction Treatment	0.059*	0.264***
	(0.027)	(0.032)
Positionality 1: Police	0.033	0.016
Investigates Crime	(0.033)	(0.037)
Positionality 2: Police	0.089**	0.018
Prosecutes Vigilantes	(0.033)	(0.038)
Crime: Cow Transport	−0.154***	
	(0.026)	
Crime: Love Jihad	0.005	
	(0.025)	
Crime: Burning Quran		0.023
		(0.030)
Crime: Tearing Down		0.045
Posters		(0.030)
Constant	−0.020	−0.168***
	(0.034)	(0.041)
Observations	5,499	4,427
R^2	0.008	0.018
Adjusted R^2	0.007	0.017
Residual Std. Error	0.997 (df = 5493)	0.992 (df = 4421)
F Statistic	8.531*** (df = 5; 5493)	15.975*** (df = 5; 4421)

Note: *p<0.05; **p<0.01; ***p<0.001

I Effect of Positionality

Our main findings demonstrate that other variables that the literature on vigilantism identifies as predictors (such as state behavior and bias) do not consistently move beliefs in the context of our experiment. In this multi-factorial framework where we are able to evaluate these variables against information about the veracity of rumors, this may suggest that correcting misinformation is *more* persuasive than information about the state's response.

These findings—departing from prior results—may, however, not hold beyond the specific multi-factorial informational environment of our design (Clayton et al 2023). Accordingly, we additionally test in this appendix whether respondents would be equally unmoved by the reported behavior of the state if they were not *also* exposed to a correction.

To do this, we rerun our main results *excluding* all the observations in which a correction appeared (about 50% of our sample).

As can be seen from tables below, in this subset of the data we find that neither Indian nor Pakistani respondents decreased their support for vigilantes when they were told that the authorities had investigated the alleged crime. However, we find that Indian respondents do, in the absence of a correction, support vigilantes to a lesser extent upon hearing that the police arrested the presumed criminals targeted by the vigilantes. This result, in conjunction with our main results on the original experiment, implies that Indian respondents use both the state's behavior towards vigilantes and the media's debunking of the rumor that prompted vigilantism as heuristics, and that the two are not independent. Interestingly, this is not the case in Pakistan where the effect of the correction is stronger than the effect of information about the state's behavior. The difference between the two countries may reflect different levels of trust in the police across the two cases.

Table I9: Main Effects of Treatments on Support for Vigilantism (No Correction)

	Support for vigilantism			
	Own Support *India*	Neighborhood Support *India*	Own Support *Pakistan*	Neighborhood Support *Pakistan*
Positionality 1: Police	−0.058	0.110	−0.075	−0.052
Investigates Crime	(0.181)	(0.174)	(0.093)	(0.095)
Positionality 2: Police	−0.396*	−0.487**	−0.156	−0.152
Prosecutes Vigilantes	(0.172)	(0.172)	(0.091)	(0.096)
Crime: Cow Transport	0.206	0.329*		
	(0.157)	(0.154)		
Crime: Love Jihad	−0.109	0.100		
	(0.153)	(0.148)		
Crime: Burning Quran			0.144	0.103
			(0.080)	(0.087)
Crime: Tearing Down			−0.050	−0.125
Posters			(0.073)	(0.081)
Constant	5.010***	5.329***	2.495***	2.806***
	(0.164)	(0.158)	(0.083)	(0.085)
Observations	2,230	2,165	2,221	2,175
R^2	0.004	0.008	0.004	0.004
Adjusted R^2	0.003	0.006	0.002	0.002
Residual Std. Error	3.376 (df = 2225)	3.287 (df = 2160)	1.724 (df = 2216)	1.826 (df = 2170)
F Statistic	2.421* (df = 4; 2225)	4.407** (df = 4; 2160)	2.053 (df = 4; 2216)	2.107 (df = 4; 2170)

Note: *p<0.05; **p<0.01; ***p<0.001

Table I10: Main Effects of Treatments on Support for Punishing Mob (No Correction)

	Should the vigilante mob be punished?	
	India	*Pakistan*
Positionality 1: Police Investigates Crime	0.104	−0.010
	(0.054)	(0.039)
Positionality 2: Police Prosecutes Vigilantes	0.216***	0.014
	(0.053)	(0.037)
Crime: Cow Transport	−0.132**	
	(0.049)	
Crime: Love Jihad	−0.001	
	(0.045)	
Crime: Burning Quran		−0.013
		(0.034)
Crime: Tearing Down Posters		0.039
		(0.032)
Constant	3.195***	3.473***
	(0.050)	(0.035)
Observations	2,770	2,179
R^2	0.009	0.001
Adjusted R^2	0.008	−0.001
Residual Std. Error	1.114 (df = 2765)	0.721 (df = 2174)
F Statistic	6.598*** (df = 4; 2765)	0.629 (df = 4; 2174)

Note: *p<0.05; **p<0.01; ***p<0.001

J Robustness Checks for Table 4

We provide a few robustness checks for results in Table 4. First we show that (null) results on the interaction hold when we change the dependent variable to punishment of the mob (Table J11). Next we switch our outgroup measure from trust in the minority to a variable that measures comfort with majority/minority friends. This is a standard affective polarization measure that asks respondents how comfortable they would be having friends from the outgroup. Tables J12 and J13 show results with comfort with the minority in each country. Across the board, we do not detect any interaction effects with the main treatment.

Table J11: Heterogeneous Effect of Trust in Minority on Support for Punishing Mob

	Should the vigilante mob be punished?	
	India	*Pakistan*
Trust in Minority	0.041	-0.138^{***}
	(0.024)	(0.024)
Correction Treatment	-0.014	0.163^{***}
	(0.065)	(0.048)
Positionality 1: Police	0.018	0.012
Investigates Crime	(0.037)	(0.024)
Positionality 2: Police	0.089^{*}	0.014
Prosecutes Vigilantes	(0.037)	(0.024)
Crime: Cow Transport	-0.184^{***}	
	(0.029)	
Crime: Love Jihad	-0.007	
	(0.028)	
Crime: Burning Quran		0.015
		(0.020)
Crime: Tearing Down		0.030
Posters		(0.019)
Correction $\times$	0.040	0.004
Trust in Minority	(0.028)	(0.028)
Constant	3.197^{***}	3.696^{***}
	(0.062)	(0.046)
Observations	5,202	4,418
R^2	0.012	0.046
Adjusted R^2	0.011	0.045
Residual Std. Error	1.089 (df = 5194)	0.640 (df = 4410)
F Statistic	9.107^{***} (df = 7; 5194)	30.619^{***} (df = 7; 4410)

Note: *p<0.05; **p<0.01; ***p<0.001

Table J12: Heterogeneous Effect of Comfort with Having Close Minority Friends on Support for Vigilantism

	Support for vigilantism			
	Own Support *India*	Neighborhood Support *India*	Own Support *Pakistan*	Neighborhood Support *Pakistan*
Comfort with Minority Friends	−0.034	−0.103	0.243***	0.313***
	(0.078)	(0.077)	(0.061)	(0.063)
Correction Treatment	0.039	−0.115	−0.486***	−0.376**
	(0.260)	(0.261)	(0.122)	(0.132)
Positionality 1: Police	0.100	0.125	−0.007	−0.064
Investigates Crime	(0.126)	(0.125)	(0.057)	(0.063)
Positionality 2: Police	−0.008	−0.119	−0.009	−0.071
Prosecutes Vigilantes	(0.125)	(0.124)	(0.059)	(0.062)
Crime: Cow Transport	0.320***	0.402***		
	(0.093)	(0.095)		
Crime: Love Jihad	−0.211*	−0.086		
	(0.094)	(0.094)		
Crime: Burning Quran			0.015	−0.036
			(0.045)	(0.050)
Crime: Tearing Down			−0.034	−0.058
Posters			(0.043)	(0.048)
Correction ×	−0.195*	−0.125	0.006	−0.019
Comfort with Minority Friends	(0.092)	(0.092)	(0.066)	(0.071)
Constant	4.951***	5.531***	2.005***	2.219***
	(0.237)	(0.231)	(0.122)	(0.126)
Observations	4,166	4,073	4,472	4,390
R^2	0.013	0.014	0.043	0.039
Adjusted R^2	0.011	0.012	0.041	0.038
Residual Std. Error	3.309 (df = 4158)	3.249 (df = 4065)	1.531 (df = 4464)	1.685 (df = 4382)
F Statistic	7.712*** (df = 7; 4158)	8.157*** (df = 7; 4065)	28.335*** (df = 7; 4464)	25.582*** (df = 7; 4382)

Note: *p<0.05; **p<0.01; ***p<0.001

Table J13: Heterogeneous Effect of Comfort with Having Close Minority Friends on Support for Punishing Mob

	Should the vigilante mob be punished?	
	India	*Pakistan*
Comfort with Minority Friends	0.093***	−0.087***
	(0.022)	(0.024)
Correction Treatment	0.052	0.146**
	(0.080)	(0.050)
Positionality 1: Police	0.032	0.012
Investigates Crime	(0.037)	(0.024)
Positionality 2: Police	0.098**	0.014
Prosecutes Vigilantes	(0.037)	(0.025)
Crime: Cow Transport	−0.152***	
	(0.029)	
Crime: Love Jihad	0.019	
	(0.028)	
Crime: Burning Quran		0.013
		(0.020)
Crime: Tearing Down		0.032
Posters		(0.019)
Correction ×	0.008	0.015
Comfort with Minority Friends	(0.027)	(0.026)
Constant	3.013***	3.623***
	(0.072)	(0.050)
Observations	5,205	4,409
R^2	0.019	0.030
Adjusted R^2	0.017	0.028
Residual Std. Error	1.081 (df = 5197)	0.645 (df = 4401)
F Statistic	14.223*** (df = 7; 5197)	19.194*** (df = 7; 4401)
Note:		*p<0.05; **p<0.01; ***p<0.001

K Ethics

The study was approved by IRB at the University of Albany, State University of New York, on May 13, 2022 (protocol number 22X134). As noted in the paper, ethical considerations were paramount in the design of our study in several ways.

First, we refrained from asking about participation in vigilantism as an outcome given the complexity of collecting information about respondent's direct involvement in (illegal, extra-judicial) violence. We decided instead to focus on support for violence, an important outcome in its own right.

Second, we worked closely with our country research teams to ensure the safety of enumerators. We opted for an audio-visual medium in part to prevent enumerators from having to read the potentially sensitive vignettes directly to respondents. The headphones also ensured privacy, particularly important in contexts which often attract the presence of third parties during interviews. Third, in order to avoid further spreading misinformation, we made sure that the focus of the vignettes was on the acts that derived from the misinformation rather than on the misinformation itself, for which we provided as few details as possible.

Next, we read a detailed debrief to respondents which contained information not only about the vignettes, but also a reminder that misinformation *in general* can lead to acts of violence and that respondents must be particularly vigilant in assessing the veracity of information. Finally, we carried out a number of pre-tests and a larger pilot prior to implementation of the survey. One of the primary goals of the pre-tests, in particular, was to ensure that the wording used was the best choice given the context.

Debriefing Script

As noted above, respondents were extensively debriefed upon finishing the interview. We reproduce below the full text of this debrief for one of our country cases (India), noting that an exactly similar - save for context-relevant details - debrief was also administered in Pakistan.

We add that these scripts were practiced during our training of enumerators and that we monitored that these were effectively delivered in the way intended.

Thank you for participating in our study. Now that you have answered our questions, we want to clarify a few important points about the materials you saw or heard.

While the scenarios you saw were inspired by real events that have happened across India in recent years, we wish to clarify that the scenarios you saw or heard were not inspired by events that happened locally, for instance in your district.

Besides, please note that neither the posts you saw nor the audio newscasts you heard were real. Instead, these were built from scratch by the research team. They are, accordingly, news items that we used for the sake of research, to help us understand how citizens respond to different types of information.

The version of these materials that you saw may additionally have been different from those that other respondents saw.

Just to be sure, let me visually explain this:

[turns the tablet towards respondent]

Note: text was in Hindi on the enumerator device.

[turns the screen towards respondent for 15 seconds so he/she has time to see it clearly]

Note: text was in Hindi on the enumerator device.

Additionally, we wish to reiterate here that too many acts of mob violence such as the ones described in these scenarios stem from false information.

We thus generally wish to encourage you to remain vigilant and to assess the material you receive on social media critically. Do not believe everything you see and consider consulting the following resources before acting on information you receive on your phone.

Many thanks for participating in the study.

Should you have questions or feel distressed by the material you saw, feel free to contact the research team at phone number xxxxxxxx.

Other Considerations

Asking questions about support for illegal, violent and polarizing behaviors in areas in which such behaviors are relatively common generates another type of risk, *on the safety of enumerators.*

Far beyond what IRBs typically recommend (insofar as they are rarely concerned with the well-being of both respondents and survey teams), we went to great lengths to ensure that the workers employed by our respective implementing partners in both countries were not at risk during the administration of the surveys.

Namely, we ensured, in collaboration with our implementing partners, that enumerator teams:

1. traveled in sufficiently large groups.
2. that they were reachable through cell phones at all times (avoiding areas insufficiently covered by cell networks).
3. had ample opportunities to report issues that may occur with certain survey items or certain questions, to both their field managers and ourselves.
4. were supervised by an experienced field manager at all times.
5. obtained, in collaboration with their field manager, formal and informal authorizations from all local stakeholder prior to fielding the survey in the localities we had randomly selected.
6. that the field manager carried at all times the contact information of local law enforcement.

They were, in addition, clearly instructed to end or interrupt, however abruptly it might be, any interview that they perceived to be causing grief, anger or other strong emotions among respondents. And they were incentivized in a way that would not penalize such behavior.

L Enumerator Fixed Effects

Table L14: Main Effects on Support for Vigilantism with Enumerator FEs

	Support for vigilantism			
	Own Support	Neighborhood Support	Own Support	Neighborhood Support
	India		*Pakistan*	
Variables				
Correction Treatment	-0.476***	-0.456***	-0.483***	-0.419***
	(0.092)	(0.094)	(0.045)	(0.049)
Positionality 1: Police	-0.020	-0.018	-0.043	-0.102*
Investigates Crime	(0.113)	(0.113)	(0.051)	(0.056)
Positionality 2: Police	-0.070	-0.228**	-0.052	-0.124**
Prosecutes Vigilantes	(0.113)	(0.115)	(0.052)	(0.056)
Crime: Cow Transport	0.334***	0.401***		
	(0.090)	(0.091)		
Crime: Love Jihad	-0.200**	-0.088		
	(0.089)	(0.090)		
Crime: Burning Quran			0.016	-0.031
			(0.045)	(0.050)
Crime: Tearing Down			-0.030	-0.054
Posters			(0.043)	(0.048)
Fixed-effects				
Enumerator ID	Yes	Yes	Yes	Yes
Fit statistics				
Observations	4,365	4,265	4,490	4,408
R^2	0.16791	0.14289	0.27579	0.22585
Within R^2	0.01127	0.01164	0.03225	0.02003

Clustered (Respondent ID) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table L15: Main Effects on Support for Punishing Mob with Enumerator FEs

	Should the vigilante mob be punished?	
	India	*Pakistan*
Variables		
Correction Treatment	0.070**	0.181***
	(0.028)	(0.020)
Positionality 1: Police	0.047	0.024
Investigates Crime	(0.035)	(0.022)
Positionality 2: Police	0.112***	0.034
Prosecutes Vigilantes	(0.035)	(0.022)
Crime: Cow Transport	-0.167***	
	(0.028)	
Crime: Love Jihad	0.005	
	(0.027)	
Crime: Burning Quran		0.013
		(0.020)
Crime: Tearing Down		0.030
Posters		(0.019)
Fixed-effects		
Enumerator ID	Yes	Yes
Fit statistics		
Observations	5,499	4,427
R^2	0.07928	0.19462
Within R^2	0.00890	0.02406

Clustered (Respondent ID) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*