Appendix B: Additional results

Research questions

We also investigated the following preregistered research questions for which we have weaker theoretical expectations. As Arnold, Reckendorf, and Wintersieck (2021) found that perceptions of accuracy differed between platforms and treatment for different partisan affiliations, we planned to investigate whether our hypotheses interact with partisanship (RQ1). We also planned to test whether the perceived accuracy of false state media tweets vary if the misinformation promotes a positive view of the country mentioned in the state media tag (RQ2). Third, we investigated whether participants who received a fact check on false tweets perceived true tweets as more accurate (an "implied truth effect;" Pennycook et al. 2020) than those who did not receive fact-checks (RQ3). Additionally, we examined how perceived accuracy changes with a country-specific state media tag relative to a generic state media tag ("state-affiliated media;" RQ4). Finally, we tested whether feelings toward the country of the state media outlet moderates the effect of the tags (RQ5).

RQ1

Table B1 reports the results of our analysis of RQ1, which sought to understand whether the treatment effects we observed differed between Democrats and Republicans. We found no evidence of a significant difference between partisan groups in this analysis, which suggests that state media tags and fact-checks have similar effects across party lines.

RQ2

Table B2 reports the results of our analysis of RQ2, a preregistered research question which asks whether the perceived accuracy of a false tweet tagged as state media will vary if the misinformation promotes a positive view of the country responsible for the state media outlet. Previous research suggests the effect of a state media tag on the perceived accuracy of a claim may vary depending on the content of the claim (Arnold et al., 2021; Nassetta & Gross, 2020). We therefore conducted a headline-level analysis testing whether the effect of state media tags on the perceived accuracy of false state media tweets varied for tweets that were about the state itself (e.g., false tweets about China attributed to Chinese state media).[1] We found no evidence of such an effect. While the baseline perceived accuracy of the country-specific tweets varied, the effects of the tags were not measurably different when the tweet content concerned the ostensible country of the state media outlet in question.

---

[1] This analysis corrects a typo in the preregistration to include indicators for the generic state media tag and fact-check label conditions.

These results suggest that respondents do not change their level of trust in or suspicion of a tweet if it seems to promote the interest of the country (i.e., by making a false claim about it). However, the tweets that reference China and Serbia did not reference the nations by name and instead relied on participants knowing that certain subregions (Xinjiang and Srebrenica, respectively) are related to them. Respondents may have been unaware of the relevance of those areas to China and Serbia, respectively, or failed to make the connection to the country in question when rating the accuracy of these claims.

Table B1. *Treatment effects on perceived accuracy of state media claims and source trust by party.*

| | Perceived accuracy | | | Source trust | |
| | False claims | | True claims | | |
	(1)	(2)	(3)	(4)	(5)
China state media tag	0.019		-0.065	-0.061	
	(0.054)		(0.053)	(0.056)	
Serbia state media tag	0.069		-0.030	-0.066	
	(0.054)		(0.053)	(0.054)	
Generic state media tag	0.043		-0.063	-0.082	
	(0.052)		(0.053)	(0.054)	
Any state media tag		0.044			-0.070
		(0.044)			(0.045)
Fact-check label	-0.194***	-0.194***	0.032	-0.065	-0.065
	(0.052)	(0.052)	(0.053)	(0.059)	(0.059)
Republican identifier/leaner	0.123**	0.123**	0.026	0.100*	0.099*
	(0.060)	(0.059)	(0.054)	(0.059)	(0.059)
China tag × Republican	0.015		0.069	0.013	
	(0.077)		(0.074)	(0.081)	
Serbia tag × Republican	-0.055		0.014	0.110	
	(0.076)		(0.075)	(0.079)	
Generic tag × Republican	-0.070		-0.020	0.058	
	(0.075)		(0.077)	(0.078)	
Any tag × Republican		-0.037			0.059
		(0.063)			(0.065)
Fact-check label × Republican	-0.026	-0.027	-0.024	0.064	0.064
	(0.075)	(0.074)	(0.074)	(0.081)	(0.081)
Controls	✓	✓	✓	✓	✓
N	2313	2313	2311	2291	2291

Note: OLS with robust standard errors; * p < .05, ** p < .01, *** p < .005 (two-sided). Perceived accuracy and source trust measured on 4-point Likert scales. Data includes partisans and leaners only. See Appendix A for stimuli and question wording.

RQ3

RQ3 asks whether we would observe evidence of an "implied truth" effect (Pennycook et al. 2020) in which participants who received a fact-check tag on false tweets would perceive true tweets as more accurate than participants who do not receive a fact-check tag on false tweets. The headline-level results, which are reported in Table B3, provide no evidence of such an effect. The estimated model includes an

indicator for being in the fact-check label condition and another for tweets seen by respondents after the first fact-check label. The latter find no measurable indication of any change in perceived accuracy.[2]

Table B2. *Treatment effects on perceived accuracy of false state media tweets.*

	Perceived accuracy
China state media tag	0.049
	(0.039)
China-related tweet	0.038*
	(0.022)
China tag × China tweet	-0.025
	(0.042)
Serbia state media tag	0.063
	(0.040)
Serbia-related tweet	-0.225***
	(0.021)
Serbia tag × Serbia tweet	-0.052
	(0.042)
Generic state media tag	0.014
	(0.036)
Fact-check label	-0.191***
	(0.036)
Controls	✓
N	7583

Note: OLS with robust standard errors clustered by respondent; * p < .05, ** p < .01, *** p < .005 (two-sided). Perceived accuracy measured on a 4-point Likert scale. See Appendix A for stimuli and question wording.

---

[2] The reported analysis represents a deviation from the preregistration, which states that the outcome variable is the perceived accuracy of true tweets seen after the first fact-check. Because this quantity is undefined for respondents not assigned to the fact-check condition, we instead conduct the analysis reported in Table B3, which also adds indicators for the generic state media tag and fact-check label conditions.

Table B3. *Fact-check label effects on perceived accuracy of true tweets.*

	Perceived accuracy
China state media tag	0.020
	(0.029)
Serbia state media tag	0.003
	(0.029)
Generic state media tag	-0.032
	(0.029)
Global Times source	-0.410***
	(0.017)
China tag $\times$ Global Times source	-0.031
	(0.030)
Serbia tag $\times$ Global Times source	-0.022
	(0.029)
Generic tag $\times$ Global Times source	-0.031
	(0.030)
Fact-check label condition	0.035
	(0.042)
After first fact-check label seen	-0.017
	(0.040)
Controls	✓
N	25291

*Note: OLS with robust standard errors (clustered by respondent for headline-level analysis); * p <.05, ** p <.01, *** p <.005 (two-sided). Perceived accuracy measured on 4-point Likert scales. See Appendix A for stimuli and question wording.*

RQ4: Specific false countries in state media tags

The results in Table 1 show no evidence of a difference in perceived false claim accuracy when a state media tag identifies a specific country rather than leaving the country in question unspecified (China: 0.026, 95% CI [-0.043, 0.094]; Serbia: 0.036, 95% CI [-0.031, 0.103]).

RQ5: Feelings toward country of state media outlet

Table B4 reports the results of a preregistered research question testing whether feelings toward the country of the state media outlet moderates the effect of the tags. We find no evidence that feelings toward either China or Serbia moderate the effect of exposure of state media tags attributing the tweets to the country in question on perceptions of the accuracy of false or true state media tweets.[3]

---

[3] This analysis corrects a typo in the preregistration to include indicators for the generic state media tag and fact-check label conditions.

Exploratory

Table B5 reports the results of our exploratory analysis testing whether pre-treatment levels of trust in state media moderate the effect of state media tags. Only 74 people (2.9%) reported a great deal of trust and confidence in state-affiliated media so we grouped these respondents with those who expressed a moderate amount (30.3%). The analysis below interacts each treatment with indicators for not very much trust and confidence and the moderate/great deal group (the omitted category as those expressing no trust and confidence at all) to avoid making a linearity assumption (Hainmueller et al., 2019).[4]

Table B4. *Treatment effects on perceived accuracy of state media tweets by country favorability.*

	False tweets	True tweets
China state media tag	0.035	-0.007
	(0.081)	(0.081)
China favorability	0.059***	0.025
	(0.018)	(0.019)
China tag × China favorability	0.001	-0.000
	(0.039)	(0.039)
Serbia state media tag	-0.027	-0.100
	(0.106)	(0.111)
Serbia favorability	-0.003	-0.038
	(0.020)	(0.022)
Serbia tag × Serbia favorability	0.032	0.035
	(0.042)	(0.045)
Generic state media tag	0.011	-0.064
	(0.036)	(0.037)
Fact-check label	-0.193***	0.027
	(0.036)	(0.037)
Controls	✓	✓
N	2533	2530

*Note: OLS with robust standard errors (clustered by respondent for headline-level analysis); * $p < .05$, ** $p < .01$, *** $p < .005$ (two-sided). Perceived accuracy measured on 4-point Likert scales. See Appendix A for stimuli and question wording.*

---

[4] Results are similar, however, if the state media trust moderator is treated as continuous (available upon request).

Table B5. *Treatment effects on perceived accuracy of state media claims by trust in state media.*

	Perceived accuracy	
	False claims	True claims
China state media tag	-0.094	-0.066
	(0.065)	(0.075)
Serbia state media tag	-0.042	-0.070
	(0.065)	(0.067)
Generic state media tag	-0.057	-0.076
	(0.069)	(0.076)
Fact-check label	-0.191***	0.024
	(0.036)	(0.037)
Not very much trust in state media	0.096*	-0.008
	(0.046)	(0.048)
Moderate/great deal of trust in state media	0.126*	0.066
	(0.054)	(0.056)
China tag × not very much	0.137	0.109
	(0.077)	(0.085)
Serbia tag × not very much	0.082	0.127
	(0.076)	(0.080)
Generic tag × not very much	0.056	0.045
	(0.078)	(0.086)
China tag × moderate/great deal	0.216**	0.027
	(0.084)	(0.091)
Serbia tag × moderate/great deal	0.167*	-0.017
	(0.083)	(0.085)
Generic tag × moderate/great deal	0.138	-0.023
	(0.086)	(0.094)
Controls	✓	✓
N	2533	2530

*Note: OLS with robust standard errors; * $p < .05$, ** $p < .01$, *** $p < .005$ (two-sided). Perceived accuracy on 4-point Likert scales. See Appendix A for stimuli and question wording.*