Supplemental Material: Does the Source of a Warning Matter? Examining the Effectiveness of Veracity Warning Labels Across Warners

Benjamin D. Horne

School of Information Sciences,
Data Science and Engineering, The Bredesen Center,
University of Tennessee Knoxville, Knoxville, TN, USA
bhorne6@utk.edu

Survey Questions and Setup

In this section, we provide the survey questions used, scales used for each question, and details about the basic flow of the experiment.

Demographic Questions:

1. What is your age?
   18 to 24
   25 to 34
   35 to 44
   45 to 54
   55 to 64
   65 or over
2. How do you describe yourself?
   Male
   Female
   Non-binary / third gender
   Prefer not to say
3. What is your racial ethnicity?
   White
   Black or African American
   Hispanic or Latino
   Native American or Alaska Native
   Asian
   Native Hawaiian or Pacific Islander
   Other
4. What is your level of education?
   Less than high school
   High school graduate
   Some college
   2-year degree
   4-year degree
   Professional degree
   Doctorate
5. What is your political leaning?
   Very Conservative
   Conservative
   Moderate
   Liberal
   Very Liberal

CRT Questions:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?
   $0.10
   $0.05
   $0.15
   $1.00
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
   100
   5
   10
   50

3. If you're running a race and you pass the person in second place, what place are you in?
   first
   second
   third
   Fourth
4. A farmer had 15 sheep and all but 8 died. How many are left?
   7
   8
   0
   6
5. Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?
   June
   Emily
   July
   March

ABI Questions:
1. The news media is capable of performing its job. (Ability)
   Completely Agree
   Somewhat Agree
   Neither Agree nor Disagree
   Somewhat Disagree
   Completely Disagree
2. The news media would not knowingly or purposely report incorrect information. (Benevolence)
   Completely Agree
   Somewhat Agree
   Neither Agree nor Disagree
   Somewhat Disagree
   Completely Disagree
3. Sound principles seem to guide the news media's behavior. (Integrity)
   Completely Agree
   Somewhat Agree
   Neither Agree nor Disagree
   Somewhat Disagree
   Completely Disagree
4. Social media companies are capable of performing their job. (Ability)
   Completely Agree
   Somewhat Agree
   Neither Agree nor Disagree
   Somewhat Disagree
   Completely Disagree
5. Social media companies would not knowingly or purposely allow incorrect information to be shared on their platforms. (Benevolence)
   Completely Agree
   Somewhat Agree
   Neither Agree nor Disagree
   Somewhat Disagree
   Completely Disagree

6. Sound principles seem to guide social media companies' behavior. (Integrity)

    Completely Agree

    Somewhat Agree

    Neither Agree nor Disagree

    Somewhat Disagree

    Completely Disagree

Experimental Conditions:

Participants are randomly assigned to one of five conditions and scroll through 8 social media posts, half of which are false. Each condition contains the same questions,, only differing in warning label design or presents. For each post participants are asked the following questions:

1. How much do you trust the information in this post?

    Distrust Completely

    Somewhat Distrust

    Neither Trust nor Distrust

    Somewhat Trust

    Trust Completely

2. Is there a particular reason you trust (or distrust) the information in this post?

3. How likely would you be to share this story on social media?

    Extremely Unlikely

    Somewhat Unlikely

    Neither Likely nor Unlikely

    Somewhat Likely

    Extremely Likely

Factual Manipulation Check

After the main discernment task, participants are asked what they thought they saw during the discernment task.

1. Did you see any warning labels on the social media posts?

    Yes

    No

2. If yes: Which of these headlines were tagged with a warning message? (Multiple selection with all headlines shown)

Debriefing

After the experiment, participants are debriefed on what information was true and false, along with details from the fact-checks. For example:

1. The following headlines were false or misleading:

    Headline 1 – Describe what was false in the headline and how we know

    Headline 2 – Describe what was false in the headline and how we know

    Headline 3 – Describe what was false in the headline and how we know

    Headline 4 – Describe what was false in the headline and how we know

Demographics

In this section, we provide additional demographic distributions that were not included in the main paper due to space. The scores created for ABI and CRT metrics are described in the main paper; however, for ease of interpretation: a higher ABI score means higher trust and a higher CRT score means higher cognitive reflection ability.

Condition	# from Facebook	# from Twitter	Total
No Label (Control)	212	200	412
Crowd	205	194	399
Platform	200	204	404
Fact Checkers	207	214	421
AI	202	211	413
Total	1,026	1,023	**2,049**

Table 1: Number of participants across each condition and platform population.

(a) Political Leaning   (b) Age   (c) Education

(d) ABI News Media   (e) ABI Social Media   (f) CRT
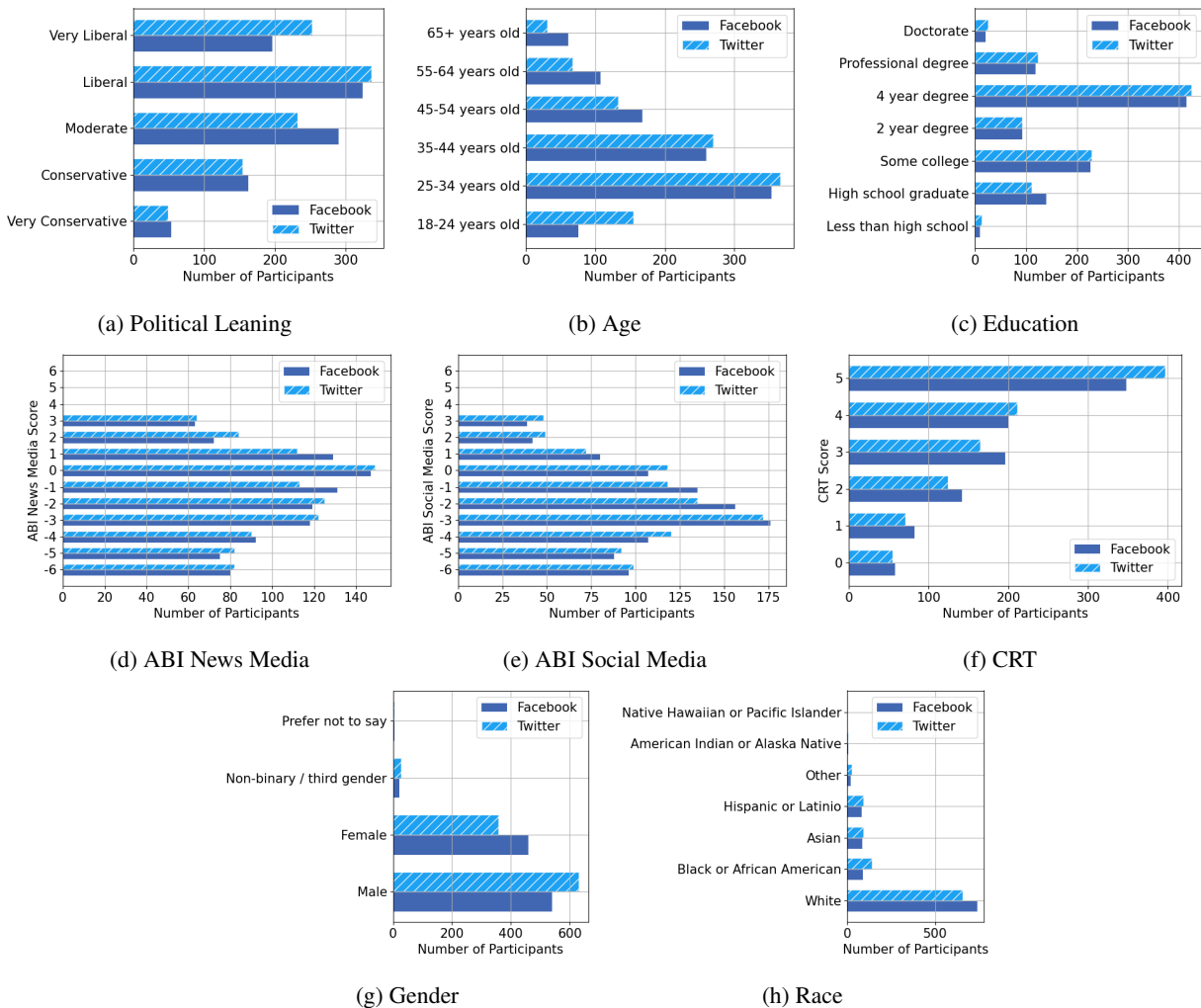
(g) Gender   (h) Race

Figure 1: Demographic data across survey participants. The sample across both the Facebook and Twitter population skews towards being more liberal, younger, and college educated.

Headlines Used and Relationship to Trust

In this section, we provide a table of the headlines that were used in the study, along with information about what organization(s) fact-checked them. Note, these headlines were not selected by topic but rather by how recently the information had been fact-check. This was done to ensure information freshness.

Post #	Headline/Post	Fact-Checked By
Post 0	Tom Hanks Arrested on 135 Counts of Child Porn Possession	Snopes and Reuters
Post 1	Ree Drummond Confirms She is 'LEAVING' The Food Network After Her Accidental 'Live' Confession On-Air	Snopes
Post 2	Google Lite: DuckDuckGo Signs Secret Deal with Bill Gates to Track Users Online	Reuters
Post 3	Chelsea Clinton: 'It's Time To Force-Jab Every Unvaccinated Child in America'	Snopes and Reuters
Post 4	"Teachers must equip children to have sexual relationships", says the United Nations	Reuters
Post 5	The Democratic Party Just Confirmed Michelle Obama Will Be Its Nominee and Nobody Noticed	FactCheck.org
Post 6	On the left is Deaon Patterson, the active shooter at large in Atlanta. On the right is Deaon Patterson on Fox News this evening.	Reuters
Post 7	DeSantis Called the National Guard to Block All Entrances to Disney World	Reuters

Table 2: False headlines

Post #	Headline/Post	Fact-Checked By
Post 8	Bucha Ukraine a year ago and today	Snopes
Post 9	NBCUniversal's Linda Yaccarino Is in Talks to Become Twitter CEO	Snopes
Post 10	The Biden administration has brought back a Trump-era immigration rule the day before Title 42 is scheduled to sunset and encourage a wave of illegal immigration into the U.S.	Snopes
Post 11	Mystery surrounds '30 foot' phallus mown into Bath's Royal Crescent days before coronation	Snopes
Post 12	Every single county in New York has experienced a federal climate disaster between 2011-2021	Politifact
Post 13	A 12-year-old is charged with murder in the shooting of a Sonic restaurant employee	Reuters
Post 14	J&J's proposed talc settlement would pay $400 million to US state AGs	Reuters
Post 15	Google to delete inactive accounts starting December	Reuters

Table 3: True headlines

Furthermore, in Figure 2, we show the relationship between individual false headlines and each condition with respect to information trust. As described in the paper, there was a significant relationship between headlines and trust across all conditions, but there was not a significant relationship between headlines and sharing intention across conditions.

Additional Sharing Analysis

In this section, we provide interaction analysis for sharing intentions that mimics the interaction analysis for trust in the paper. Specifically, in Table 4 we show the OLS interaction analysis for ABINews, ABISocial, and PoliticalLeaning. In Figure 3 we show the interaction plots for ABINews, ABISocial, and PoliticalLeaning.

(a) *ABINews Model*	coef	std err	P> |t|
Intercept	-3.1949	0.057	0.000***
Platform	-0.3218	0.081	0.000***
Crowd	-0.3088	0.082	0.000***
AI	-0.3981	0.080	0.000***
FC	-0.3093	0.081	0.000***
ABINews	0.0729	0.018	0.000***
ABINews X Platform	-0.0855	0.026	0.001**
ABINews X Crowd	-0.0637	0.026	0.015*
ABINews X AI	-0.0403	0.026	0.126
ABINews X FC	-0.0667	0.025	0.009*

(b) *ABISocial Model*	coef	std err	P> |t|
Intercept	-2.9529	0.065	0.000***
Platform	-0.4407	0.091	0.000***
Crowd	-0.4613	0.092	0.000***
AI	-0.5693	0.088	0.000***
FC	-0.4160	0.091	0.000***
ABISocial	0.1632	0.019	0.000***
ABISocial X Platform	-0.1051	0.028	0.000***
ABISocial X Crowd	-0.1068	0.028	0.000***
ABISocial X AI	-0.0960	0.027	0.000***
ABISocial X FC	-0.0875	0.02	0.001**

(c) *PoliticalLeaning Model*	coef	std err	P> |t|
Intercept	-2.6102	0.136	0.000***
Platform	-0.4718	0.185	0.011*
Crowd	-0.7034	0.191	0.000***
AI	-0.6898	0.187	0.000***
FC	-0.4898	0.186	0.008**
PoliticalLeaning	-0.2564	0.049	0.000***
PoliticalLeaning X Platform	-0.0858	0.067	0.199
PoliticalLeaning X Crowd	0.1766	0.069	0.011*
PoliticalLeaning X AI	0.1246	0.068	0.068
PoliticalLeaning X FC	0.0946	0.066	0.155

Table 4: OLS regressions on the participant-level sharing intentions for (a) ABINews, (b) ABISocial, and (c) PoliticalLeaning. Significance codes are: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Other Mixed Effects Models

In Tables 4 through 7, we show the results of our Mixed Linear models with differing condition references (rather than in reference to control). These results reflect what was shown in the paper using the much simpler Tukey's HSD test, found in Table 1 of the paper. The Mixed Linear model implementation we used in this work comes from the statsmodel Python3 package, version 0.14.0. The basic formulation of the model using this package is:

```
model = smf.mixedlm("TrustScore ~ C(Condition, Treatment(reference='NoLabel')) +
    C(PlatformGroup) + Age + Education + PoliticalLeaning + CRT + ABINews + ABISocial", df,
    groups="ResponseID").fit()
```

where df is a Pandas Dataframe that only contains data from interaction with false headlines and $ResponseID$ is a label for each participant (Remember, we are grouping repeated measures). Note that despite this being Python code, statsmodel uses the R-style formulation. Each variable in the model corresponds to a column in a CSV file.

Table 5: Mixed Linear Model Regression Results (False in reference to platform)

Model:	MixedLM	Dependent Variable:	TrustScore
No. Observations:	8141	Method:	REML
No. Groups:	2041	Scale:	0.1646
Min. group size:	2	Log-Likelihood:	-5312.5823
Max. group size:	4	Converged:	Yes
Mean group size:	4.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	-0.380	0.036	-10.561	0.000	-0.451	-0.310
C(Condition, Treatment(reference='Platform'))[T.AI]	-0.051	0.024	-2.105	0.035	-0.098	-0.003
C(Condition, Treatment(reference='Platform'))[T.Crowd]	-0.008	0.024	-0.327	0.744	-0.055	0.040
C(Condition, Treatment(reference='Platform'))[T.Fact Checkers]	-0.019	0.024	-0.784	0.433	-0.066	0.028
C(Condition, Treatment(reference='Platform'))[T.No Label]	0.109	0.024	4.547	0.000	0.062	0.156
C(PlatformGroup)[T.TW]	0.008	0.015	0.512	0.609	-0.022	0.038
Age	-0.008	0.006	-1.358	0.174	-0.020	0.004
Education	-0.015	0.006	-2.627	0.009	-0.026	-0.004
PoliticalLeaning	-0.047	0.007	-6.627	0.000	-0.061	-0.033
CRT	-0.025	0.005	-4.851	0.000	-0.035	-0.015
ABINews	-0.013	0.003	-3.842	0.000	-0.020	-0.007
ABISocial	0.022	0.004	5.790	0.000	0.014	0.029
ResponseID Var	0.076	0.010				

Table 6: Mixed Linear Model Regression Results (False in reference to crowd)

Model:	MixedLM	Dependent Variable:	TrustScore
No. Observations:	8141	Method:	REML
No. Groups:	2041	Scale:	0.1646
Min. group size:	2	Log-Likelihood:	-5312.5823
Max. group size:	4	Converged:	Yes
Mean group size:	4.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	-0.388	0.036	-10.680	0.000	-0.459	-0.317
C(Condition, Treatment(reference='Crowd'))[T.AI]	-0.043	0.024	-1.768	0.077	-0.090	0.005
C(Condition, Treatment(reference='Crowd'))[T.Fact Checkers]	-0.011	0.024	-0.451	0.652	-0.058	0.036
C(Condition, Treatment(reference='Crowd'))[T.No Label]	0.117	0.024	4.857	0.000	0.070	0.164
C(Condition, Treatment(reference='Crowd'))[T.Platform]	0.008	0.024	0.327	0.744	-0.040	0.055
C(PlatformGroup)[T.TW]	0.008	0.015	0.512	0.609	-0.022	0.038
Age	-0.008	0.006	-1.358	0.174	-0.020	0.004
Education	-0.015	0.006	-2.627	0.009	-0.026	-0.004
PoliticalLeaning	-0.047	0.007	-6.627	0.000	-0.061	-0.033
CRT	-0.025	0.005	-4.851	0.000	-0.035	-0.015
ABINews	-0.013	0.003	-3.842	0.000	-0.020	-0.007
ABISocial	0.022	0.004	5.790	0.000	0.014	0.029
ResponseID Var	0.076	0.010				

Table 7: Mixed Linear Model Regression Results (False in reference to AI)

Model:	MixedLM	Dependent Variable:	TrustScore
No. Observations:	8141	Method:	REML
No. Groups:	2041	Scale:	0.1646
Min. group size:	2	Log-Likelihood:	-5312.5823
Max. group size:	4	Converged:	Yes
Mean group size:	4.0		

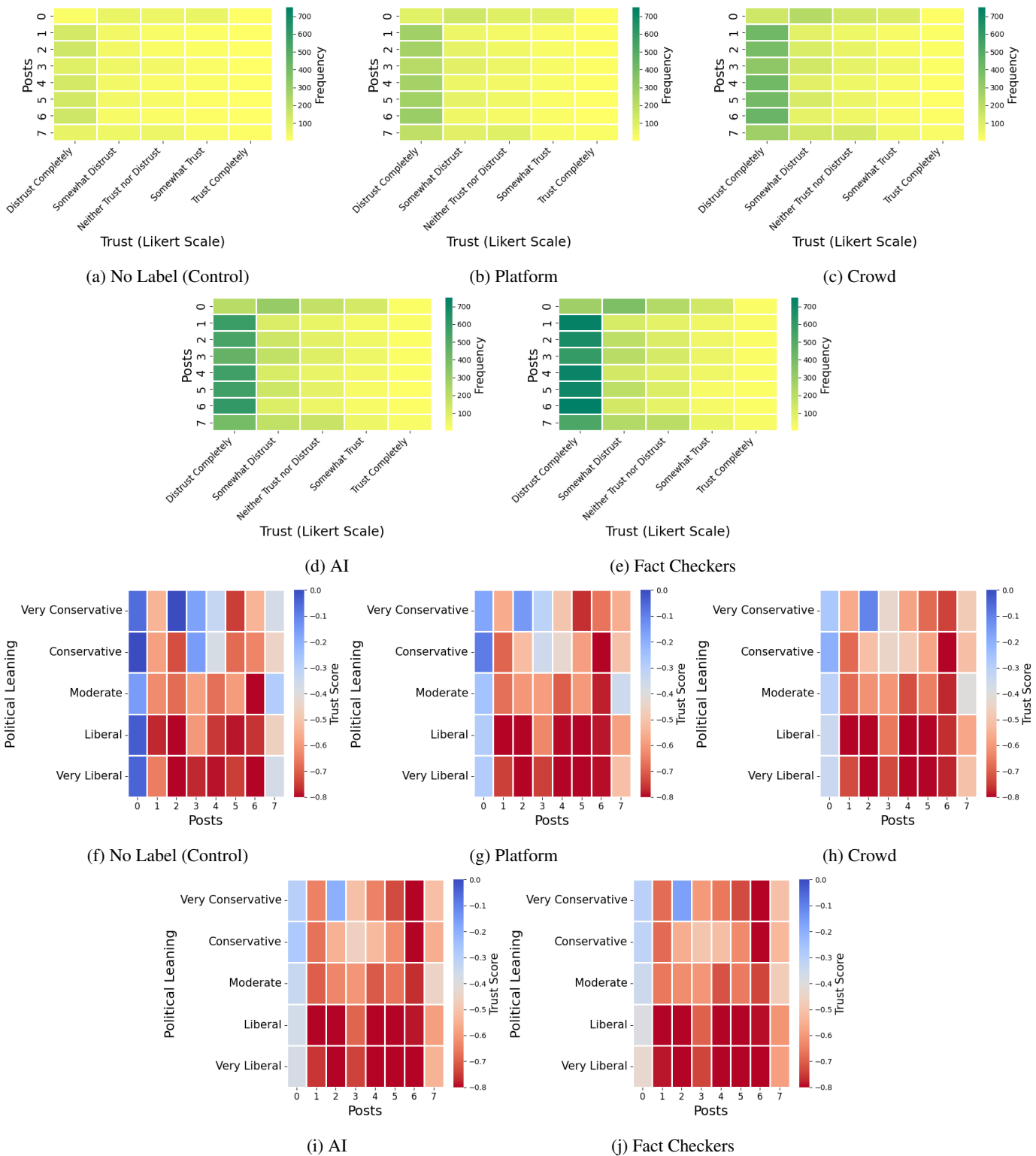|	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	-0.431	0.036	-12.029	0.000	-0.501	-0.361
C(Condition, Treatment(reference='AI'))[T.Crowd]	0.043	0.024	1.768	0.077	-0.005	0.090
C(Condition, Treatment(reference='AI'))[T.Fact Checkers]	0.032	0.024	1.339	0.181	-0.015	0.079
C(Condition, Treatment(reference='AI'))[T.No Label]	0.160	0.024	6.673	0.000	0.113	0.207
C(Condition, Treatment(reference='AI'))[T.Platform]	0.051	0.024	2.105	0.035	0.003	0.098
C(PlatformGroup)[T.TW]	0.008	0.015	0.512	0.609	-0.022	0.038
Age	-0.008	0.006	-1.358	0.174	-0.020	0.004
Education	-0.015	0.006	-2.627	0.009	-0.026	-0.004
PoliticalLeaning	-0.047	0.007	-6.627	0.000	-0.061	-0.033
CRT	-0.025	0.005	-4.851	0.000	-0.035	-0.015
ABINews	-0.013	0.003	-3.842	0.000	-0.020	-0.007
ABISocial	0.022	0.004	5.790	0.000	0.014	0.029
ResponseID Var	0.076	0.010				

Figure 2: (Top) Frequency of trust (Likert scale) per false post, where darker (more green) is more frequent. In each condition, a Chi-Squared Test of Independence showed a significant relationship between the post and frequency of trust ($p < 0.000$). However, as our Mixed Effects model indicated, warning label conditions each decrease trust in false posts. (Bottom) Average trust scores per political leaning per false post. While information was not always political, some posts were trusted more or less by differing political groups. However, the heatmaps mostly reflect the findings from the Mixed Effects model in the paper, conservatives trusted false information more than liberals.
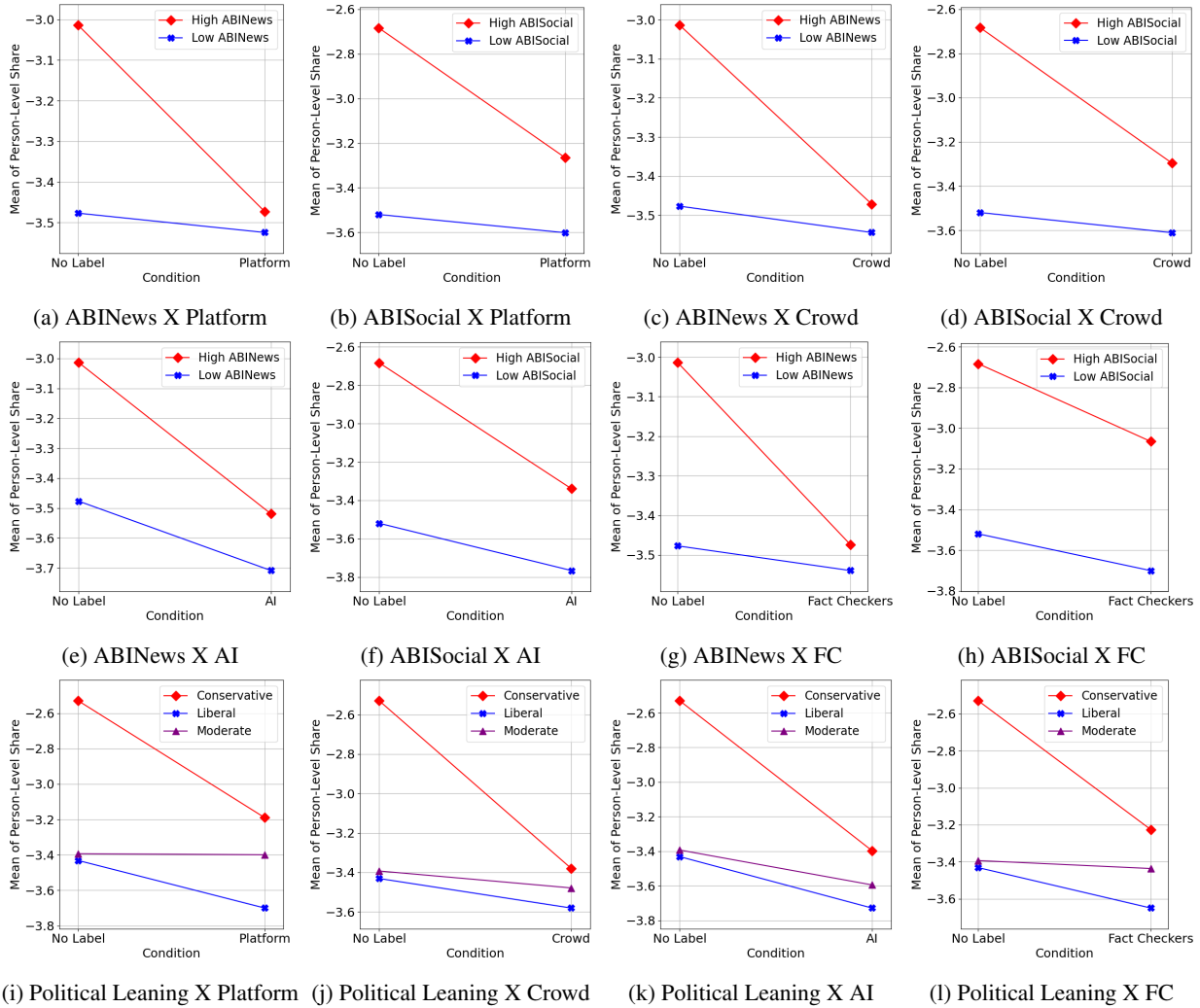
Figure 3: Interaction plots of participant-level sharing intentions across high and low ABINews and ABISocial subgroups, and political leaning groups. Significance can be found in Table 4.
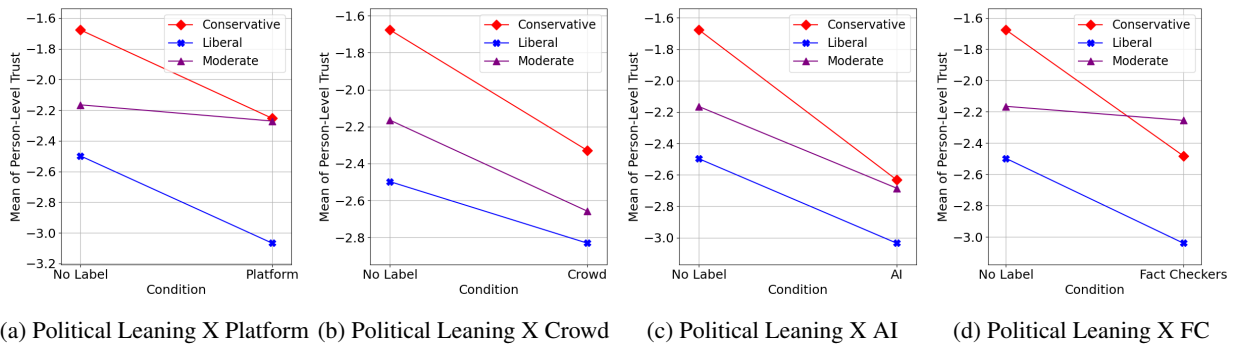

Figure 4: Non-significant interaction plots of participant-level trust across political leaning groups.

Table 8: Mixed Linear Model Regression Results (False in reference to Fact Checkers)

Model:	MixedLM	Dependent Variable:	TrustScore
No. Observations:	8141	Method:	REML
No. Groups:	2041	Scale:	0.1646
Min. group size:	2	Log-Likelihood:	-5312.5823
Max. group size:	4	Converged:	Yes
Mean group size:	4.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	-0.399	0.036	-11.043	0.000	-0.470	-0.328
C(Condition, Treatment(reference='Fact Checkers'))[T.AI]	-0.032	0.024	-1.339	0.181	-0.079	0.015
C(Condition, Treatment(reference='Fact Checkers'))[T.Crowd]	0.011	0.024	0.451	0.652	-0.036	0.058
C(Condition, Treatment(reference='Fact Checkers'))[T.No Label]	0.128	0.024	5.378	0.000	0.081	0.175
C(Condition, Treatment(reference='Fact Checkers'))[T.Platform]	0.019	0.024	0.784	0.433	-0.028	0.066
C(PlatformGroup)[T.TW]	0.008	0.015	0.512	0.609	-0.022	0.038
Age	-0.008	0.006	-1.358	0.174	-0.020	0.004
Education	-0.015	0.006	-2.627	0.009	-0.026	-0.004
PoliticalLeaning	-0.047	0.007	-6.627	0.000	-0.061	-0.033
CRT	-0.025	0.005	-4.851	0.000	-0.035	-0.015
ABINews	-0.013	0.003	-3.842	0.000	-0.020	-0.007
ABISocial	0.022	0.004	5.790	0.000	0.014	0.029
ResponseID Var	0.076	0.010				