THE EFFECT OF FEEDBACK ON NEWS-VERIFICATION DEMAND: EXPERIMENTAL EXPERIMENTAL EVIDENCE 8th AnnualTexas Experimental Association Symposium

Dario Trujano-Ochoa and Jose Gloria

Texas A&M

DO YOU THINK THIS IS TRUE?



Would you verify this headline?



How well can Americans distinguish real news headlines from fake ones?

 $\label{eq:FIGURE 1: https://today.yougov.com/politics/articles/45855-americans-distinguish-real-fake-news-headline-poll} FIGURE 1: https://today.yougov.com/politics/articles/45855-americans-distinguish-real-fake-news-headline-poll$

"Overall, Americans correctly classify a headline about two-thirds of the time."

"... similar accuracy overall for the real and fake headlines; the average for each set is about 7 out of 10."

- Does feedback affect the demand for verification and accuracy?
 - Individual accuracy rates
 - Others accuracy rates
- Do political variables change the demand for verification and accuracy?
 - Headline's political content
 - Participants political position

- The spread of misinformation through digital platforms is faster and deeper than real information (Vosoughi, Roy and Aral, 2018).
 - Mostly by people (not robots) who share inadvertently (Arin, Mazrekaj and Thum, 2023).
- Fact-checking, and labeling content are relevant tools for countering the effects of misinformation. (Kozyreva et al., 2024; Bateman and Jackson, 2024).
 - Point out false information.
- Previous interventions focus on increasing ability and attention in sharing accurate information (Pennycook et al. (2021) and Pennycook and Rand (2022)).

- Direct measure of the demand for verification.
 - Classification-verification game with actual headlines.
 - BDM mechanism
- Evaluation of the effects of feedback on the demand for verifying headlines.
- Analyze the differences between political and non-political headlines.
- Increases the research on misinformation in Mexico

EXPERIMENTAL PARAMETERS

In the experiment we control variables that are important in the verification decision:

- Decision without signal
 - Initial classification $c \in \{t, f\}$, about
 - State of the world $\omega \in \{T, F\}$.
- P(T) = P(F) = 0.5
- $U_T = U_F = \pi = 10MXN$
- $U_{TF} = U_{FT} = 0$
- Perfect Signal

•
$$P(s = t|T) = P(s = f|F) = 1$$



- Mexico City, Summer 2024
- 195 undergrad students
 - UNAM (largest and most important university in Mexico)
 - Psychology
 - Mathematics
 - IPN (second most important public university in Mexico)
 - Informatics
- 42.05% men
- 20.1 years old

Rounds and Blocks

Headline Number 18

Time left to complete this page: 0:01

Please classify the following headline: (If your classification is correct, you could earn an extra 10 MXN.)

Iran Censored the Olympics; All Women Appear with Rectangles or Asterisks Covering Them

Your Classification:

O The information is accurate O Contains false information

How much are you willing to pay to verify this news?



Block	Rou	unds
1	10 headlines	Non-political
2	10 headlines	Non-political
3	10 headlines	Non-political
4	10 headlines	Political
5	10 headlines	Political

- For each headline
 - Classification: $c \in \{t, f\}$
 - Verification: $WTP(c) \in [0, 5]$
 - BDM mechanism
 - 20 second limit to answer both questions. Time distribution
 - At least 75% of the headlines classified:
 - 2.5% participants excluded

TABLE 1: Feedback Treatments

Treatment Group	Feedback at the End of the
	Block
Control Group	No feedback on accuracy was
	given.
Individual Feedback	Personal accuracy rate for the
	block conditional on the head-
	lines participants classified as ac -
	curate or fake.
Others Feedback	Average accuracy rate of other
	participants conditional on the
	headlines others classified as ac -
	curate or fake.

RESULTS

Variable	Control	Individual	Others
Age	20	20.1	20.2
Male	0.516	0.319	0.435
Support Gov	0.203	0.232	0.161
Oppose Gov	0.156	0.203	0.194
Missing Headlines	0.033	0.026	0.055
Accuracy Estimate	0.574	0.525	0.542
Accuracy Estimate Others	0.526	0.509	0.495
Accuracy	0.618	0.603	0.594
Classification $(c = a)$	0.492	0.505	0.508
WTP	2.81	2.65	2.46
N Participants	64	69	62

FEEDBACK ABOUT OTHERS



FIGURE 2: Empirical CDF of the willingness to pay by treatment. To create this graph, the average WTP per block was calculated.

	Dependent variable:			
=	WTP	Accuracy	WTP Pol.	Accuracy Pol.
	(1)	(2)	(3)	(4)
Individual Feedback	$ \begin{array}{c} -0.192 \\ (0.212) \end{array} $	-0.014 (0.014)	-0.209 (0.228)	-0.017 (0.017)
Others Feedback	-0.353^{*}	-0.010	-0.410^{*}	-0.034^{*}
Round	(0.205) 0.002 (0.002)	$(0.014) \\ 0.0002 \\ (0.001)$	$(0.224) \\ 0.004 \\ (0.004)$	$(0.019) \\ -0.012^{***} \\ (0.001)$
'True' (c = t)	0.210^{***} (0.054)	0.002 (0.004)	0.203*** (0.065)	0.213^{***} (0.006)
Political	0.137 ^{**} (0.054)	0.037^{*} (0.022)		
Gov. Supporter	0.475^{**}	0.040^{***}	0.403	0.013
Favor Gov. News	(0.211)	(0.010)	(0.243) -0.039 (0.043)	-0.207^{***} (0.016)
Gov. Critic	0.215 (0.229)	0.004 (0.014)	(0.010) (0.202) (0.247)	-0.066^{***} (0.023)
Supporter X Favor	()	()	0.071	0.007
Critic X Favor			0.086 (0.077)	(0.0302) (0.098^{***}) (0.031)

Note:

p < 0.1; p < 0.05; p < 0.05; p < 0.01

The SE clustered at the individual level and first block excluded

CONCLUSIONS

- Providing Feedback on the probabilities that other people correctly classify headlines could backfire.
 - People receiving feedback on the accuracy or others demanded less verification.
- There is no evidence of Overconfidence in the experiment.
- People are more willing to pay for headlines when
 - They believe are true
 - The headlines are political
 - Government supporters
- The probability of classifying a headline as accurate decreases if a headline favors the government and the participant is against the government.

THANK YOU



dariotrujanoochoa@ucsb.edu https://dariotrujanoochoa.github.io/

APPENDICES

LATENCY BY CLASSIFICATION



FIGURE 3: Frequency of time spent on headlines classified as true and false.

TIME AND ACCURACY



FIGURE 4: Average time spent on each headline and the proportion of correct classifications against the average classification as "true."

CONFIDENCE ELICITATION • BACK TO BLOCK

Confidence in Block Classification 3

Answer the following questions with the probability in percentage terms. Where 100 means the event always occurs, 0 means it never occurs, and 50 means it occurs half of the time.

Please consider the block of 10 news headlines that you just classified:

You classified 5 headlines as "The information is accurate" and 5 as "Contains false information".

One of ti What is f	ne 5 h he pro	eadlines you classified as accurate will be selected at random. bability that the headline is actually accurate?
70	~	
Dne of ti What is f	ne 5 h he pro	eadlines you classified as false will be selected at random. bability that the headline is actually false?
55	~	
Now, o headl	cons nes:	ider the classification that other participants made in this block of 10 news
A headlii What is f	ne clas he pro	ssified as accurate by another participant will be selected at random. obability that the headline is actually accurate?
50	~	

A headline classified as false by another participant will be selected at random. What is the probability that the headline is actually false?



Next

FIGURE 5: Screenshot of the translated Confidence elicitation as seen by the participants.

VERIFICATION'S PROBLEM (DECISION MAKING FRAMEWORK)

EXPECTED UTILITIES FROM THE INITIAL CLASSIFICATION

- Decision without signal
 - Initial classification $c \in \{t, f\}$, about
 - State of the world $\omega \in \{T, F\}$.

$$EU_{\text{no signal}}(t) = P(T|t) \cdot U_T + P(F|t) \cdot U_{TF}$$
$$EU_{\text{no signal}}(f) = P(F|f) \cdot U_F + P(T|f) \cdot U_{FT}$$

$$P(s = f|F), P(s = t|T) > 0.5$$

• Conditional expected utilities

 $EU_{\text{new classification}}(s=t,c) = P(T|s=t,c) \cdot U_T + P(F|s=t,c) \cdot U_{TF}$

 $EU_{\text{new classification}}(s = f, c) = P(F|s = f, c) \cdot U_F + P(T|s = f, c) \cdot U_{FT}$

VALUE OF VERIFICATION WITH FOLLOWING THE SIGNAL

• Expected value of following the signal

$$EU_{\text{signal}}^{\text{update}}(c) = P(s = t|c) \cdot EU_{\text{new classification}}(s = t, c) + P(s = f|c) \cdot EU_{\text{new classification}}(s = f, c)$$

• Value of the signal

$$V(c) = EU_{\text{signal}}^{\text{update}}(c) - EU_{\text{no signal}}(c)$$

• Back to Experimental Parameters

HEADLINES SELECTION

- 60 headlines tested on Prolific
 - Accurate headlines: NewsGPT
 - All fact checked
 - Fact-checked headlines: AnimalPolitico and Verificado
 - Popular and false headlines
- 50 selected for the Experiment
 - Similar accuracy rates in each block

▲ Back to Headlines