The Effect of Feedback on News-Verification Demand: Experimental Evidence*
(Check the latest version here.)

By Dario Trujano-Ochoa† and Jose Gloria‡

This study is the first to elicit the value individuals place on verifying real headlines, focusing on the role of feedback in shaping verification behavior and confidence. In a lab experiment, 184 participants classified headlines as accurate or fake, reporting their willingness to pay (WTP) for a perfect signal under three feedback conditions: control, individual, and group. Results show that explicit feedback on others' accuracy reduces the perceived value of verification, potentially discouraging fact-checking. The findings highlight how feedback influences verification demand. By addressing gaps in existing literature and focusing on Mexico, this study provides novel insights into the design of interventions to combat misinformation.
JEL: C93, D83, D91
Keywords: Misinformation, Feedback, Verification, Willingness to Pay

When an individual encounters a headline, they face a decision problem: should they invest time and effort to verify its accuracy before forming an opinion or acting on it? This decision often hinges on the perceived cost of verification versus the potential consequences of relying on misinformation. If individuals are informed about how difficult it is to identify false claim they may reassess the value of verification. Knowing the challenge may increase their willingness to verify, as they recognize the higher risk of error, or conversely, it could discourage them if they perceive the task as too daunting, especially if the headline aligns with their pre-existing beliefs.

The spread of misinformation through digital platforms is faster and more profound than factual information (Vosoughi et al., 2018), and it is primarily people (not bots) who share misinformation inadvertently (Arin et al., 2023). To avoid making decisions based on false assumptions it is essential to focus on interven-

* The authors want to thank the professors who discussed the early versions of the present project: Gary Charness, Cesi Cruz, Daniel Martin, Ignacio Esponda, Ryan Oprea, Sevgi Yuksel, Erik Eyster, and to the specialists and professors in Mexico who increased the discussion and understanding of the relevant problem of misinformation: Grisel Salazar, Daniel Moreno, Horacio Larreguy, Antonio Arechar, Pablo Soto, and Arturo Bouzas. We also thank AlianzaMX for the fellowship that allowed the authors to travel to Mexico to develop this research.

† UCSB, USA, dariotrujanoochoa@ucsb.edu.

‡ UCLA, USA, josegloria@ucla.edu.

tions at the individual decision-making level. Misinformation undermines the public's ability to make informed decisions, erodes trust in institutions, and fuels polarization, which threatens democratic societies.

Most existing research focuses on the effects of exogenous interventions and overlooks the intrinsic value individuals place on verifying the accuracy of the information they receive. This value ultimately determines the time and effort they are willing to invest in checking the veracity of headlines. In this paper, we refer to verification as searching for further information to reveal if a statement is accurate or false.[1] Individual verification is essential to fight misinformation, increasing the effectiveness of measures that point out false statements like debunking and fact-checking.[2] Thus, verification complements other efforts to fight misinformation.[3]

Overconfidence has been identified as a factor contributing to the spread of misinformation, as overconfident individuals are less likely to invest time and effort in verifying the accuracy of information. However, evidence supporting this relationship primarily comes from survey data. In this paper, we formalize this reasoning with an explicit framework for the decision problem of information classification, where individuals highly confident in their ability to identify false information are predicted to rationally reduce their demand for verification.

Evidence suggests that feedback reduces overconfidence, which motivated us to explore its potential effects in this study. However, the primary objective was to assess how a feedback intervention impacts the value individuals place on verification. While interventions like debunking, inoculation, and digital literacy provide indirect feedback on accuracy when classifying headlines, this research directly measures the demand for verification in a lab experiment, examining how explicit feedback alters this valuation.

This is the first study, to our knowledge, that elicits the value of verifying real headlines. It contributes to the literature by examining the impact of feedback on both confidence levels and the perceived value of verification, providing evidence that feedback on the accuracy of other participants can reduce this value. By focusing on headlines from Mexico—the largest Spanish-speaking country where misinformation and polarization have recently increased—this research addresses a critical gap in the literature, which has predominantly concentrated on interventions studied in the U.S. and Europe (Kozyreva et al., 2024; Bateman and Jackson, 2024). The study's experimental design controls for participant hetero-

[1]Not every statement and information can be classified in this way. Normative statements (e.g., "Headlines should prioritize impactful language to capture readers' attention.") are examples. This paper focuses on headlines easily classified as accurate or false.

[2]In this paper, we distinguish fact-checking from verification by defining fact-checking as a policy aimed at identifying and disseminating information about fake news, whereas verification involves an individual's proactive choice to seek additional details.

[3]For example, Facebook (Meta) collaborates with fact-checking organizations to tag false information (https://transparency.meta.com/policies/community-standards/misinformation), and X uses CommunityNotes (https://x.com/communitynotes?lang=en) to let the users decide through an algorithm over their ratings.

geneity, ensuring that its results are broadly applicable to diverse populations. By moving beyond the typical focus on exogenous interventions, this paper innovatively explores how feedback mechanisms influence individuals' active demand for verification, offering insights into how such mechanisms can enhance the effectiveness of fact-checking in combating misinformation.

I. Previous Literature

People's overconfidence has been shown to impact their ability to discern fake news, leading to greater engagement with false information. According to Lyons et al. (2021), individuals tend to overestimate their ability to distinguish between real and fake news. Overconfident individuals are not only more likely to believe in and share fake news but are also unaware of their limitations in identifying misinformation. This overconfidence exacerbates the spread of false information, especially when the misinformation aligns with their political or ideological beliefs. Similarly, Pennycook and Rand (2020) found that overconfidence in one's cognitive abilities correlates with a higher likelihood of accepting false claims as true. Ortoleva and Snowberg (2015) found that overconfidence due to correlation neglect leads to higher polarization. Together, these studies highlight overconfidence's role in the dissemination of fake news by decreasing verification and increasing misinformation sharing.

Encouraging fact-checking and verification behaviors has been effective in combating misinformation. Reviews by Kozyreva et al. (2024) and Bateman and Jackson (2024) noted that promoting media literacy, fact-checking, and labeling content are vital tools to counter misinformation. However, Aslett et al. (2024) found that verification tools, like search engines, could backfire by amplifying misinformation when individuals are overconfident in their search skills. This research indicates that while verification behaviors are crucial, the limitations of tools like search engines must be considered in curbing misinformation effectively.

Other interventions, such as accuracy prompts, encourage users to pause and assess content accuracy before sharing. For instance, Pennycook et al. (2021) introduced an "accuracy nudge" intervention that reduced fake news spread by prompting users to reflect on accuracy, highlighting the positive impact of slight cognitive adjustments on online behavior. Similarly, Pennycook and Rand (2022) focused on interventions promoting cognitive engagement to reduce misinformation dissemination, demonstrating that simple adjustments in thought processes significantly improve the quality of shared content.

The literature shows mixed results on feedback's effects on overconfidence. While some incentivized studies have found feedback reduces overconfidence (Ferraro, 2005; Eberlein et al., 2011; Kogelnik, 2022), others in educational contexts report no effects (Pulford and Colman, 1997; Erat et al., 2022). Moreover, some studies reveal feedback's asymmetric effects on motivated beliefs. Oprea and Yuksel (2022) observed that feedback increases subjective probabilities of outperforming others, while Thaler (2024) noted that individuals update beliefs asymmetrically when feedback reinforces ego-related worldviews. Kartal and Tyran (2022) showed that overconfident participants continue voting despite low-accuracy signals, though they did not directly measure demand for information. Additionally, Moore and Healy (2008) distinguished between overestimation, overplacement, and overprecision, noting that feedback often has minimal effects on recalibrating overconfidence, particularly for difficult tasks where individuals tend to overestimate their performance.

This research aims to establish if feedback impacts the demand for verification. Unlike prior studies focusing on the accuracy of misinformation distinction or verification strategies, this study offers experimental evidence on feedback's causal effect on verification demand—a novel contribution to misinformation literature.

II. Hypotheses on the Effects of Feedback

This study is structured around three key hypotheses that explore the impact of feedback on willingness to pay (WTP) to verify the accuracy of the headlines and the confidence. Under the assumption of rationality, participants' WTP should be proportional to the value they give to information. These hypotheses are integrated into the methodology to test their validity in a controlled environment, using neutral headlines to minimize the effects of motivated reasoning.

HYPOTHESIS 1: *Participants are generally overconfident and will expect better performance in classifying headlines than what is reflected in the feedback they receive.*

Following the literature on overconfidence, this hypothesis suggests that participants tend to overestimate their classification accuracy before receiving feedback. In this experiment, participants' predictions about their classification accuracy are expected to be higher than the accuracy indicated by their actual performance, as revealed by the feedback.

HYPOTHESIS 2: *Participants' willingness to pay (WTP) for verification will be higher when they receive feedback on group classification accuracy than when they receive personal performance feedback.*

According to previous results on the asymmetric effect of feedback, participants who receive feedback on the performance of others could perceive this feedback as more informative and objective. As a result, they will place a larger value on the verification process and be willing to pay more to ensure the accuracy of the headlines they classify. The expectation is that group feedback, being less affected by motivated reasoning, will lead to a higher WTP as participants seek to mitigate the perceived difficulty of the task.

HYPOTHESIS 3: *Participants' willingness to pay (WTP) for verification is influenced by the political content of the headlines, with differing effects depending on whether the headline favors or opposes the current government.*

1) *Supporters of the Government: Lower WTP for verification of favorable headlines; higher WTP for verification of unfavorable headlines.*

2) *Opponents of the Government: Higher WTP for verification of favorable headlines; lower WTP for verification of unfavorable headlines.*

When participants are presented with headlines that contain political content, it is hypothesized that their WTP for verification will vary depending on whether the headline aligns with their political beliefs. Specifically, if a headline is favorable to the current government, participants who support the government will likely have lower WTP for verification. This is because they are more inclined to accept information that aligns with their pre-existing beliefs without seeking further verification. Conversely, participants who oppose the current government may exhibit higher WTP for verification of favorable headlines, as they may be more skeptical of information that contradicts their beliefs and, thus, more motivated to confirm its accuracy.

On the other hand, for headlines unfavorable to the current government, supporters of the government may demonstrate higher WTP for verification, driven by a desire to challenge or disprove information that opposes their political views. Opponents of the government, however, may show lower WTP for verification of unfavorable headlines, as they may be more likely to accept information that aligns with their negative views of the government without the need for additional confirmation.

III. Decision-Making in the Classification of Headline Accuracy

This section describes the agent's decision-making problem of classifying, verifying, and reclassifying a headline as accurate or fake. Classifying is a signal detection problem, and purchasing additional signals on this decision requires calculating the expected value of sample information (EVSI). This instrumental value of information is assumed to be equal to the willingness to pay for verification. The preset setting provides a framework for understanding the decision-making process when deciding whether to verify headlines.

A. Problem Setup Without Purchasing a Signal

Consider an agent tasked with classifying headlines as accurate (a) or fake (f) ($c \in \{a, f\}$). The state of the world is $\omega \in \Omega = \{A, F\}$, with the prior probability of encountering a fake headline denoted by $P(\omega = F) = p_f$.[4] Consequently, the prior probability of encountering an accurate headline is $1 - p_f$.

The agent's utility for correctly classifying a headline as accurate is U_A, and for correctly classifying a headline as fake is U_F;. Conversely, the utility is $U_{AF} < U_A$ for misclassifying a fake headline as accurate and $U_{FA} < U_F$ for misclassifying

[4]We simplify $P(\omega = F)$ to $P(F)$. For $c, s \in \Omega$, we specify.

an accurate headline as fake. The condition is case-insensitive for evaluating the correct classification (i.e., $c = \omega$ means a correct classification).

The probability of classifying correctly the headline is determined by $P(c = a|A)$ and $P(c = f|F)$ with $1 < \frac{P(c=a|A)}{P(c=a|F)}$ and $1 < \frac{P(c=f|F)}{P(c=f|A)}$ to assure that the initial classification c is informative in the sense that the initial classification c gives information relative to the prior probability of each state ω.[5] This is stated formally in the following proposition.

PROPOSITION 1: *Informativeness of the initial classification c.*
$1 < \frac{P(c=\omega|\omega)}{P(c=\omega|\tilde{\omega}\neq\omega)}$ *if and only if* $P(\omega|c \neq \omega) < P(\omega) < P(\omega|c = \omega)$

Notice that commonly found assumptions $0.5 < P(s = a|A) = q_a$ and $0.5 < P(s = f|F) = q_f$ are sufficient to make the signal S informative according to proposition 1. The proof of this proposition can be found in the appendices.

The expected utilities when a headline is classified as accurate, $EU_{\text{no signal}}(a)$, and as fake, $EU_{\text{no signal}}(f)$, are given respectively by the equations:

$$EU_{\text{no signal}}(a) = P(A|a) \cdot U_A + P(F|a) \cdot U_{AF}$$

$$EU_{\text{no signal}}(f) = P(F|f) \cdot U_F + P(A|f) \cdot U_{FA}$$

From the previous equations, it is clear that if the headline's classification is informative, reading a headline is valuable because the expected utility is larger than considering the prior probabilities alone.

B. *Conditional WTP Analysis*

This section will show the optimal willingness to pay (WTP) for signal S after the initial classification. The WTP is the maximum amount that an agent would pay to observe signal S.

The decision to purchase information happens after observing a headline once the agent has classified the signal. Therefore, the value of the signal depends on c. We are assuming that sequential information acquisition is optimal. The problem of sequential decision-making was stated in general by Wald (1947), and Arrow et al. (1949) analyzed how to learn from sequential information.

This section presents the condition that makes verifying the initial classification valuable. After initially classifying the headline as accurate ($c = a$) or false ($c = f$), the agent can reclassify the headline $r \in \{a, f\}$ based on the signal's realization $s \in \{a, f\}$. Let's consider first a valuable signal S with the conditions in proposition 2.

[5] We are assuming here that the classification is a signal to the same agent without considering the content of a headline h which is most likely multidimensional. This classification process also follows an optimization process where $c = \omega \iff \frac{P(\omega|h)}{P(\omega|\tilde{\omega}\neq\omega|h)} > \frac{U_A - U_{FA}}{U_F - U_{AF}}$. However, following the objectives of the present research, we focus on analyzing the informativeness of the initial classification c in proposition 1 without analyzing the properties of the headlines or the payoffs.

DEFINITION 1: *A signal is valuable if* $EU(r = s) \geq EU(r = c)$

The informativeness of the signal is determined by $P(s = a|A) = q_a$. We need a strong enough signal S so that the signal is valuable and the optimal decision to reclassify is to follow the signal ($r = s \in \{a, f\}$). Also, we assume that the signal realization s is independent of the previous classification c conditional on the stare of the world $\omega \in \{A, F\}$ (i.e. $P(s|\omega, c) = P(s|\omega)$).

PROPOSITION 2: *Conditions for Valuable Signal*
A signal S is valuable if and only if

$$\frac{P(\omega|s = \omega, c \neq \omega)}{P(\tilde{\omega}|s = \omega, c \neq \omega)} < \frac{U_\omega - U_{\tilde{\omega}\omega}}{U_{\tilde{\omega}} - U_{\omega\tilde{\omega}}} \equiv U_\omega$$

with $\tilde{\omega} \in \Omega, \tilde{\omega} \neq \omega$.

We are also assuming that the initial classification is valuable and therefore follows the analogous condition $\frac{P(\omega|c=\omega)}{P(\tilde{\omega}|c=\omega)} < U_\omega$.

By allowing the agent to update their classification based on the signal, we account for the dynamic decision-making process. The WTP to verify is derived by comparing the expected utility with the signal (considering reclassification) to the expected utility without the signal. This approach shows the impact of additional information on improving decision-making accuracy. The detailed mathematical steps and proofs are provided in the appendix. The expected utility of reclassification is calculated by updating the agent's posterior beliefs using Bayes' rule and comparing the expected utilities with and without reclassification.

For an agent tasked with classifying headlines as accurate or fake, a signal S indicating the state of the world must be sufficiently strong to ensure that the agent reclassifies based on this signal.

WTP Equation

Initially, the agent classifies a headline as either accurate (a) or fake (f). The agent updates their beliefs upon receiving a signal s, which can either confirm or contradict the initial classification. The posterior probabilities are calculated using Bayes' rule. For example, the posterior probability of the headline being accurate given the signal $s = a$ and the initial classification c is:

$$P(A|s = a, c) = \frac{q_a \cdot P(A|c)}{q_a \cdot P(A|c) + (1 - q_f) \cdot P(F|c)}$$

Similarly, the posterior probability of the headline being fake given the signal $s = f$ and the initial classification c is:

$$P(F|s = f, c) = \frac{q_f \cdot P(F|c)}{(1 - q_a) \cdot P(A|c) + q_f \cdot P(F|c)}$$

The agent's decision to reclassify based on the signal depends on the expected utilities. The expected utility of reclassification given the signal $s = a$, or $s = f$, are respectively:

$$EU_{\text{new classification}}(s = a, c) = P(A|s = a, c) \cdot U_A + P(F|s = a, c) \cdot U_{AF}$$

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

The combined expected utility of updating the signal, considering both possible signals, is:

$$\begin{aligned} EU^{\text{update}}_{\text{signal}}(c) = \quad & P(s = a|c) \cdot EU_{\text{new classification}}(s = a, c) + \\ & P(s = f|c) \cdot EU_{\text{new classification}}(s = f, c) \\ = \quad & [q_a \cdot P(A|c) + (1 - q_f) \cdot P(F|c)] \cdot [P(A|s = a, c) \cdot U_A + P(F|s = a, c) \cdot U_{AF}] + \\ & [(1 - q_a) \cdot P(A|c) + q_f \cdot P(F|c)] \cdot [P(F|s = f, c) \cdot U_F + P(A|s = f, c) \cdot U_{FA}] \end{aligned}$$

The WTP to verify the headline is calculated by comparing the expected utility with the signal to the expected utility without the signal:

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

C. *Simplifying Assumptions*

Let's assume equal prior probabilities $p_f = 0.5$ and equal utilities $U_A = U_F = 1$ and $U_{AF} = U_{FA} = 0$. Also, assume that the prevalence of fake and accurate news is the same $P(A) = P(F) = p_f = 0.5$. These assumptions about the payoffs allow us to interpret the signal's value purely in probability terms related to its informativeness.

Substituting these assumptions into the expected utility equations, we get:

$$EU_{\text{no signal}}(a) = P(A|a) \cdot 1 + P(F|a) \cdot 0 = P(A|a) = \frac{P(a|A)}{P(a|A) + P(a|F)}$$

$$EU_{\text{no signal}}(f) = P(F|f) \cdot 1 + P(A|f) \cdot 0 = P(F|f) = \frac{P(f|F)}{P(f|A) + P(f|F)}$$

And the expected utilities simplify to:

$$EU_{\text{new classification}}(s = a, c) = P(A|s = a, c)$$
$$EU_{\text{new classification}}(s = f, c) = P(F|s = f, c)$$

Finally, the combined expected utility of updating the signal, considering both possible signals, is:

$$
\begin{aligned}
EU^{\text{update}}_{\text{signal}}(c) = \quad & [q_a \cdot P(A|c) + (1-q_f) \cdot P(F|c)] \cdot P(A|s=a,c) + \\
& [(1-q_a) \cdot P(A|c) + q_f \cdot P(F|c)] \cdot P(F|s=f,c)
\end{aligned}
$$

Perfect Signal

Here, we calculate the WTP considering the condition $q_f = q_a = 1$; perfect signal. This assumption ensures that the signal is strong enough to follow even without the other simplifying assumptions and substantially simplifies the interpretation of $EVSI(c)$. For the case $c = f$ and $s = a$: $q_a \cdot P(A|f) > (1-q_f) \cdot P(F|f) \iff P(A|f) > 0$. The case $c = s$ and $s = f$ requires $P(F|a) > 0$. Both conditions are satisfied by the construction of the problem.

This assumption simplifies the expected utility of observing the signal S. Thus, the combined expected utility with the signal is:

$$EU^{\text{update}}_{\text{signal}}(c) = P(A|c) \cdot 1 + P(F|c) \cdot 1 = P(A|c) + P(F|c) = 1$$

Therefore,

$$V(c) = \begin{cases} 1 - P(A|a), c = a \\ 1 - P(F|f), c = f \end{cases} \tag{1}$$

The value of the signal S is the difference between the posterior probability of reclassifying correctly after observing the signal and the posterior probability of initially classifying correctly.

Notice that if we change the payoff of a correct answer such that $U_A = U_F > U_{AF} = U_{FA}$, we only have to multiply the posterior probabilities difference by $\pi = U_A - U_{AF}$ to get $V(c)$. Therefore, the willingness to pay to verify should be:

$$WTP(c) = \pi V(c) \tag{2}$$

The WTP decision is based solely on accuracy probability.

IV. Experimental Design: Classification-Verification Game

A. Overview

This experiment tests whether different types of feedback influence participants' accuracy in classifying information and their willingness to pay (WTP) for verification. Participants are tasked with categorizing headlines as accurate or fake and reporting their WTP for verification. They evaluated 50 headlines in five blocks, with feedback provided in one of three experimental conditions: control (no feedback), individual feedback, and group feedback. The design measures participants' classification accuracy, confidence, and willingness to pay for verification. All this is framed through the decision-making problem presented in section III. Following the main task, participants complete a survey on demographics and political orientation.

B. *Experimental Blocks and Feedback Treatments*

Participants completed five blocks designed to measure classification accuracy, confidence, and demand for verification through WTP. In each block, participants received ten headlines in random order, which they were instructed to classify as either *accurate* (a) or *fake* (f), with an equal prior probability ($P(A) = P(F) = 0.5$) of each state. For each correctly classified headline, participants earned a utility of 10 Mexican Pesos (MXN), regardless of whether the classification was *accurate* or *fake* ($U_A = U_F = 10$ MXN). Conversely, misclassifications, whether mistakenly classifying an accurate headline as *fake* or a fake headline as *accurate*, yielded a utility of 0 MXN ($U_{AF} = U_{FA} = 0$ MXN). The classification and WTP for each headline decision had a 20-second time limit. This utility structure incentivized participants to classify accurately.

For each headline, participants also indicated their willingness to pay (WTP) to access a perfect signal (S) that could reveal the headline's actual status. The perfect signal, available for purchase, would reveal the actual state of each headline with certainty ($P(s = a|A) = q_a = 1$ and $P(s = f|F) = q_f = 1$).

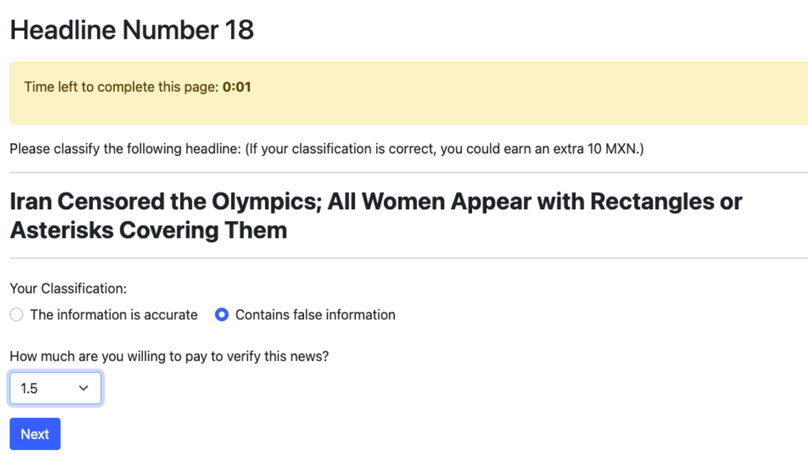


FIGURE 1. SCREENSHOT OF THE TRANSLATED CLASSIFICATION-VERIFICATION GAME AS SEEN BY THE PARTICIPANTS.

After completing all classifications and WTP decisions within a block, participants reported their estimated probability of correctly classifying the headlines by themselves and others. There was no time limit when participants reported their confidence. This provided a self-assessed confidence measure for the block.

After estimating their probabilities at the end of each block, participants received feedback according to the treatment group to which they were randomly assigned. Feedback treatments were designed to inform participants about their classification performance, either individually or relative to others. The control group was used as a reference. The feedback types and their descriptions are presented in 1. By block, all treatment groups were shown a summary of the times they classified a headline as *accurate* or *fake*, and the feedback treatments were shown the accuracy rates conditional on the headlines classified as *accurate* or *fake*.

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 headlines participants classified as *accurate* or *fake*.
Others Feedback	Average accuracy rate of other participants conditional on the headlines others classified as *accurate* or *fake*.

This structure allowed researchers to observe how different feedback types influenced participants' accuracy, confidence, and valuation of the verification signal throughout the experiment.

C. *Experimental Procedures*

Sample Information

The participants were 184 undergraduate students in Mexico. The average age was 20 years old, and 55% of them were women. They were recruited from UNAM (National Autonomous University of Mexico) and IPN (National Polytechnic Institute), the first and second most important public schools in Mexico[6]. The experiment occurred at the schools where participants were studying in September 2024.

Incentives and Utility

Participants receive a utility of 10 Mexican Pesos (MXN) for each correctly classified headline, whether it is accurate or fake ($U_A = U_F = 10$ MXN). Con-

[6]In the national ranking, UNAM is the most important university, and IPN can be ranked third (https://www.usnews.com/education/best-global-universities/mexico) or forth (https://www.topuniversities.com/university-rankings-articles/world-university-rankings/best-universities-mexico), depending of the ranking.

versely, they receive a utility of 0 MXN for misclassifications, whether they mistakenly classify an accurate headline as fake or a fake headline as accurate ($U_{AF} = U_{FA} = 0$ MXN). This setup incentivizes participants to classify accurately and to value the signal appropriately based on its accuracy-assurance potential.

Post-Experiment Survey and Payment

After all blocks, participants complete a survey collecting demographic data and assessing their support for the current government. For payment, one block is randomly selected at the experiment's end, and participants receive 10 MXN for each correctly classified headline in the chosen block, thus linking final earnings directly to classification accuracy.

D. Methodological Considerations

This experiment controls for variables that are relevant in the field to focus purely on the effects of feedback. The believed probability of receiving fake news, the interest in classifying correctly, and the verification quality change the value of verifying. The key parameters and assumptions—such as the equal prior probabilities, equal utilities, and a perfect signal—simplify the problem and allow for a focus on the probabilistic aspects of classification and verification.

Willingness to Pay for Verification and Confidence

We used the BDM mechanism to measure participants' WTP for verification. This was presented as a second price auction against a computer randomly choosing numbers from 0 to 5. Participants also chose a number between 0 and 5 from a drop-down menu to participate in the auction. If they win the auction, the signal is verified, and they win 10 MXN independently of the classification they made of the headline. If they lose the auction by betting less than the computer, their payoff will depend only on their initial classification.

In the field, verifying is a discrete decision based on each headline's expected gains and costs. However, the willingness to pay to verify directly measures the expected gains hidden in a discrete decision. Two people verifying (or not) can have different values for the information. Eliciting the WTP can allow a continuous measure of the value of verification and, therefore, its demand and how this might change depending on treatment.

We used the method in (Wilson and Vespa, 2018) to measure the confidence levels to present the binarized scoring rule in plain text. To increase this measure's reliability, we implemented this simple description of the problem (Charness et al., 2021), and assuring the participants that it is in their best interest to report their true beliefs (Danz et al., 2022) was implemented. The exact wording of this elicitation can be found in figure C1 of the appendix.

Headlines Selection

The online publication AnimalPolitico[7] and VerificadoMX[8] were used as the sources to find relevant fake news circulating in Mexico. These are the most relevant fact-checking efforts recognized in Mexico. To find headlines that were real but difficult to classify, the authors used NewsGPT[9]. The authors verified these headlines independently. All the headlines generated were verified independently by the authors. The authors selected 60 headlines from these sources: 30 political and 30 non-political, half of which were true and the other half false. Also, from the political headlines, 15 were classified as information that favored the government, and 15 opposed the government.

To select the 50 headlines for the final experiment and the order in which the blocks were placed, we ran a study in Prolific among Mexicans whose first language was Spanish. We asked for the classification of the headlines with the same incentives as in the final experiment and measured the probability of each headline being classified correctly. The headline composition of the blocks was made so that they have similar difficulty levels. The list of headlines used in the experiment can be found in the appendix table C.C2.

Using neutral headlines minimizes the potential influence of motivated reasoning, allowing the experiment to focus on the effects of feedback and overconfidence. At the same time, political news can show the impact of motivated reasoning on the demand for information.

Political position

In the exit survey, participants had to answer some questions about politics and their answers to the questions: 1) who did you vote for? and 2) if the approved work made by AMLO[10] as president. If a participant answered "Morena"[11] to the first question and "Agree" to the second, that person was classified as a supporter of the government. If a participant voted for any other party and disagreed with the statement, they were classified as opposing the government. The polarization in Mexico has increased just as in the US. However, the main division is in options about AMLO and the political party he founded to run for president (Morena)[12].

[7] https://animalpolitico.com/verificacion-de-hechos

[8] https://verificado.com.mx/

[9] The request was made in August, around three weeks before the start of the first session: https://chatgpt.com/g/g-NnU2wmnZ5-newsgpt-chat-with-hundreds-of-news-sources/c/7e750031- b534-481c-83cf-2dc6917d98b4

[10] This is a common way that people and media use to refer to Andres Manuel Lopez Obrador. The president of Mexico from 2018 to 2024.

[11] This is the name of the incumbent party during the federal elections in 2024.

[12] https://apnews.com/article/mexico-election-polarized-divided-heat-violence-4d5f620f0f8f9b7ef6efa8b3083561a8

This is an opinion shared in traditional media[13], academics and journalists[14]. The experiment happened in September 2024, more than three months after the Presidential election in Mexico.

V. Results

Table 2 shows the average values of the most important variables grouped by treatment. The first four variables are participants' characteristics; ages didn't change much between treatments, and the proportion of female participants was slightly larger in the Individual group, but in general, the groups were balanced. The proportion of participants supporting the government ranged from 16.4% to 24%, while the proportion of participants opposing the government ranged from 14.5% to 21.8%. Also, the proportion of missing headlines due to not answering in the 20-second limit was always less than 5%. Regarding confidence, the average probability of making the right decision is always higher for themselves than others. Also, their estimated accuracy rate is below the accuracy rate observed ("Correct"). The results in the accuracy and correctness suggest underestimation and overplacement (following Moore and Healy (2008)'s classification).

Table 2—Summary by treatment of the main variables. The average (proportion) of each variable is presented for each treatment.

Variable	Control	Individual	Others
Age	19.9	20	20.4
Female	0.452	0.672	0.509
Support Gov	0.21	0.239	0.164
Oppose Gov	0.145	0.209	0.218
Missing Headlines	0.029	0.023	0.048
Accuracy Estimate	0.57	0.522	0.548
Accuracy Estimate Others	0.522	0.505	0.494
Correct	0.62	0.606	0.596
Classification (c = a)	0.497	0.507	0.511
WTP	2.82	2.64	2.42
N Participants	62	67	55

TABLE 3—REGRESSION ON THE CONFIDENCE AND WILLINGNESS TO PAY. THE REGRESSION IS ON THE DATA AT THE LEVEL OF ROUNDS EXTRAPOLATING THE BLOCK CONFIDENCE TO THE CONFIDENCE IN EACH HEADLINE.

	Dependent variable:	
	Confidence	WTP
	(1)	(2)
Individual Feedback	−4.705	−0.211
	(3.027)	(0.215)
Others Feedback	−2.537	-0.385^{*}
	(3.237)	(0.209)
Block	-1.077^{*}	−0.012
	(0.552)	(0.023)
'Accurate' (c = a)	1.389	0.236^{***}
	(0.956)	(0.057)
Correct	−0.032	−0.039
	(0.449)	(0.042)
Age	0.977	−0.035
	(0.610)	(0.052)
Male	3.353	0.019
	(2.594)	(0.178)
Confidence		0.001
		(0.003)
Political	5.510^{***}	0.194^{***}
	(1.457)	(0.056)
Support Gov	8.006^{**}	0.503^{**}
	(3.329)	(0.205)
Against Gov	5.028	0.251
	(3.265)	(0.238)
Constant	33.673^{***}	3.198^{***}
	(12.440)	(1.049)
Observations	8,903	8,903
R^2	0.053	0.034
Adjusted R^2	0.052	0.032
Note:	*p<0.1; **p<0.05; ***p<0.01	

A. *Confidence*

This section it is summarized the results from the regression 1 in the table 3. We explained the levels of confidence (measured by the reported $P(A|c = a)$ and $p = P(F|c = f))$ by the treatments, participants characteritics and headline propierties.

After making their classifications, participants are asked to report the probability that they believe their classifications are correct. This self-reported probability allows for the calculation of overconfidence metrics. Overconfidence is assessed by comparing the participants' reported probabilities of correct classification with the actual probabilities derived from the task. Specifically, overconfidence for accurate classifications is calculated as $O_a = \frac{P(A|a)}{\hat{P}(A|a)}$, and for fake classifications as $O_f = \frac{P(F|f)}{\hat{P}(F|f)}$. An overall measure of overconfidence is also calculated as $O = \frac{P(c=\omega)}{\hat{P}(c=\omega)}$, where c is the classification and ω is the true state of the world.

Higher Levels of Confidence

Participants demonstrated higher confidence levels when classifying **political headlines**, with a notable increase among **government supporters**. This trend indicates that confidence is context-sensitive, with political relevance amplifying participants' certainty in their classifications. Supporters of the government, in particular, tended to exhibit a stronger conviction that their classifications were accurate, especially for politically aligned headlines.

Confidence Decreases with Experience

An analysis across blocks reveals that **confidence declined as participants gained experience** with the task. This reduction in confidence suggests that familiarity with the task did not reinforce belief in accuracy; instead, participants became more cautious, possibly recognizing the complexity or ambiguity in classifying certain headlines. This pattern is consistent with learning effects observed in prior experimental studies, where experience moderates initial overconfidence.

No Evidence of Feedback Effects on Confidence

Our results indicate **no statistically significant impact of feedback**—whether individual or group—on participants' reported confidence in their classifications. This outcome suggests that feedback in this context did not affect participants'

[13] https://apnews.com/article/mexico-election-polarized-divided-heat-violence-4d5f620f0f8f9b7ef6efa8b3083561a8

[14] https://www.eluniversal.com.mx/tendencias/la-reflexion-de-denise-maerker-sobre-las-elecciones-2024-que-se-volvio-viral/

accuracy self-assessment, even though feedback types varied across conditions. This finding aligns with other studies showing limited feedback effects on confidence when feedback does not address specific decision outcomes.

Lack of Average Overconfidence

Contrary to the hypothesized overconfidence in headline classification, **participants were not generally overconfident**. Average confidence was in line with observed accuracy rates, indicating a realistic self-assessment. This aligns with prior findings in similar tasks where participants are directly incentivized for accuracy and can adjust beliefs based on task difficulty.

B. *Willingness to Pay (WTP) for Verification*

This sections shows the results on the willingness to pay (WTP). This results can be seen in table 3.

Group Accuracy Feedback Reduces WTP

Receiving **group accuracy feedback** significantly lowered participants' WTP to verify classifications. This reduction in demand for verification suggests that collective feedback may diminish the perceived need for individual verification, possibly due to an implicit assumption of greater task simplicity or shared accuracy. This outcome underscores the role of social feedback in shaping verification behavior in information classification tasks. This result can also be observed in figure 2 where the distribution of WTP for the control group second order stocastically domintes the distribution of WTP in the "Others" treatmen.

Higher WTP for Verification of Believed-True Information

Participants exhibited a greater **willingness to pay (WTP) for verification of information they classified as true**. This indicates a verification bias, whereby participants seek to confirm rather than challenge initial beliefs. This finding is consistent with behavioral patterns in decision-making under uncertainty, where individuals are more likely to invest in information that reinforces their prior beliefs. This result can also be observed in figure 3 where the distribution of WTP for the *false* classification second order stocastically domintes the distribution of WTP in *accurate* classification. This is true for every treatment.

Increased WTP for Political News Verification

Participants demonstrated an **increased WTP for verifying political news**, emphasizing the perceived importance of accuracy for politically sensitive content. This effect was amplified among **government supporters**, who displayed a higher demand for verification, potentially reflecting a higher stake in the perceived accuracy of politically favorable information. This finding suggests that

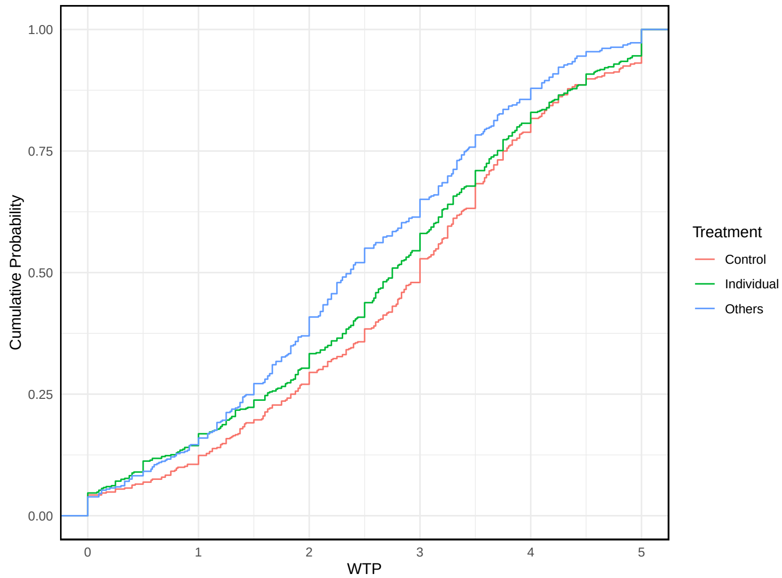


Figure 2. Empirical CDF of the willingness to pay by treatment. To create this graph the average WTP per block was calculated.

political alignment influences not only classification behavior but also the demand for verification.

C. *Effects on Political Headlines*

In this section, we analyze only the headlines that contain political content and can be classified as "Favorable" or "Unfavorablet" to the Government. The results can be observed in the regressions displayed in table 4.

WTP Patterns in Political News

The WTP results observed across all headlines were preserved when analyzing political news exclusively. However, in contrast to general findings, **political alignment did not significantly affect WTP** within this subset, indicating that demand for verification in politically charged content may be driven by factors other than simple partisan alignment.

Classification Patterns

Further analysis reveals that participants were **more likely to classify a headline as accurate** if it favored the government, a trend especially pronounced

Table 4—Regression on the Willingness To Pay and Classification. The regression is on the data at the level of rounds extrapolating the block confidence to the confidence in each headline.

	Dependent variable:	
	WTP	Accurate (c = a)
	(1)	(2)
Individual Feedback	−0.214	0.015
	(0.240)	(0.022)
Others Feedback	−0.450*	0.021
	(0.241)	(0.024)
Block	0.042	−0.003
	(0.050)	(0.015)
Confidence	0.001	−0.0001
	(0.004)	(0.001)
'Accurate' (c = a)	0.242***	
	(0.072)	
Correct	−0.040	0.228***
	(0.054)	(0.007)
Age	−0.048	0.006
	(0.056)	(0.005)
Male	0.057	−0.010
	(0.202)	(0.019)
Support Gov	0.392	−0.033
	(0.255)	(0.032)
News Favor Gov	−0.033	0.159***
	(0.045)	(0.025)
Against Gov	0.254	0.170***
	(0.261)	(0.034)
Support Gov X Favor Gov	0.092	0.064
	(0.082)	(0.043)
Favor Gov X Against Gov	0.083	−0.261***
	(0.079)	(0.050)
Constant	3.420***	0.165
	(1.160)	(0.107)
Observations	3,603	3,603
R^2	0.033	0.069
Adjusted R^2	0.029	0.066
Note:	*p<0.1; **p<0.05; ***p<0.01	

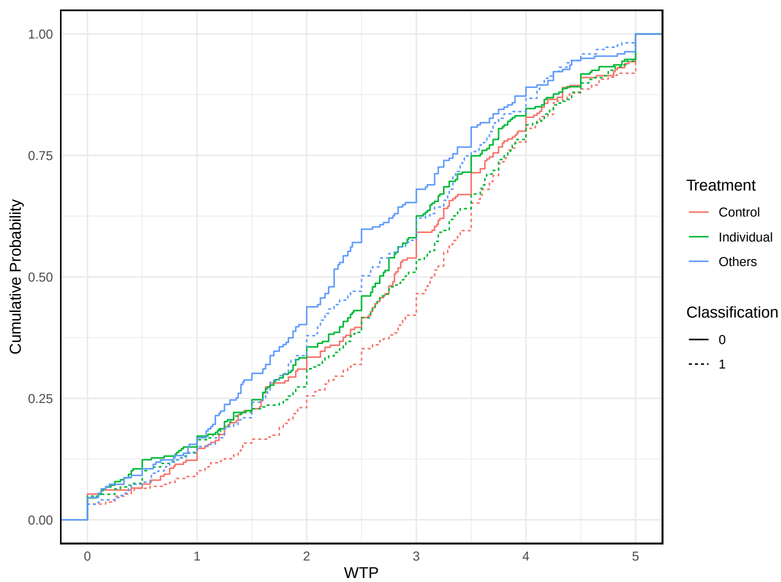


FIGURE 3. EMPIRICAL CDF OF THE WILLINGNESS TO PAY BY TREATMENT AND CLASSIFICATION OF THE HEADLINE. TO CREATE THIS GRAPH THE AVERAGE WTP PER BLOCK WAS CALCULATED.

among those who opposed the government. This suggests that partisan perception plays a role in classification tendencies, with individuals more likely to label favorable headlines as accurate even when they hold opposing political views. Additionally, for headlines correctly classified initially, participants were more inclined to classify them as accurate, reinforcing the tendency to affirm initial judgments.

VI. Discussion

We found a negative relationship between the feedback on other's performance and the willingness to pay for information. Considering that underconfidence was a more prevalent trait of the participants, this result follows the theory; they could be learning that the task is easier than expected, and they will consider that the probability of correctly classifying is higher, which reduces the value of new verifying the headline. However, we didn't find a significant effect of feedback on the probability they thought they got a correct classification (level of confidence). The willingness to pay was asked by each headline (50 observations), while the confidence was asked at the end of the block. This mismatch creates a noisy relation between these variables since the average confidence is not linked to individual headlines, where participants can be very confident or very unsure about the veracity of the headline. We believe that this noise in the measure of both variables hides the relationship between the expected probability of correct classification and willingness to pay to verify.

A. Perfect Verification

The methodology presented here simplifies the world in a way that allows us to explore the causal effects of feedback on the value of verification. In the field, we have verification practices that could be imperfect and different utilities for believing fake news and rejecting accurate information. Providing feedback in the real world could change the importance of each kind of mistake. This could be behind the effects of highlighting the importance of accuracy observed in Pennycook et al. (2021) and Pennycook and Rand (2022).

The best way to verify is an open question not addressed in the current research; the question is about overconfidence as a mechanism behind low levels of verification. We present a perfect signal to avoid motivated misinterpretation of the signal's likelihoods (Thaler (2024)) and also to avoid the discussion of the real information value of a specific verification practice. The results from Thaler (2024) are very important since a no-completely-informative signal will allow a biased updating belief that also decreases the WTP for the signal. Even online searching of news, one of the most common practices promoted in digital literacy programs, could backfire (Aslett et al. (2024); Hoes et al. (2023)). Also, having imperfect signals might make purchasing any information suboptimal if participants expect their original classification to remain the same even after the

B. Classification and Verification Behavior

The results indicate that participants classified headlines as accurate ($c = a$) approximately 50% of the time ($P(c = a) = 0.5$), aligning with the known prior probabilities of each state being equal ($P(A) = P(F) = 0.5$), which they were informed about. In terms of confidence, there was a slight asymmetry: $P(A|c = a) = 54.7\%$ and $P(F|c = f) = 53.6\%$. This difference was minimal, suggesting that participants may be adhering to the martingale property, as their classification accuracy remained consistent with the prior.

Interestingly, participants showed a greater willingness to pay (WTP) for verification when they classified a headline as accurate ($c = a$), suggesting that they were more motivated to confirm their accurate classifications. This finding implies an inequality that can be expressed through the decision-making analysis:

$$
\begin{aligned}
& WTP(c=a) > WTP(c=f) \\
& \Leftrightarrow \\
& 1 - P(A|c=a) > 1 - P(F|c=f \\
& \Leftrightarrow \\
& P(A|c=a) < P(F|c=f)
\end{aligned}
$$

Additionally, participants were more accurate when classifying headlines as accurate: $P(c = A|c = a) = 67.3\%$ compared to $P(c = F|c = f) = 60.2\%$. This contradicts the general expectation that people would exhibit more caution in marking information as accurate. One possible explanation is that participants may have an internalized utility structure that penalizes certain types of misclassifications, such as inaccurately labeling false information as accurate (e.g., $U_{AF} < 0$). However, if such a utility structure were in effect, we might expect a lower overall proportion of headlines classified as accurate due to increased caution.

These findings highlight subtle asymmetries in confidence and verification demand that could inform future research on verification behavior, especially regarding how individuals internalize error costs in classification tasks.

VII. Conclusion

This study provides experimental evidence on the effects of feedback in a news classification task, examining its impact on participants' willingness to pay (WTP) for verification and their confidence in their classifications. Our findings have several implications for understanding how individuals value verification under different feedback conditions, particularly in a context where misinformation is prevalent.

One key finding is the impact of group feedback on verification demand. Contrary to the hypothesis that feedback would motivate fact-checking, we observed that providing feedback on group accuracy actually reduced participants' WTP for verification. This result suggests that group feedback may inadvertently signal that personal verification is less critical, thereby reducing individual motivation for fact-checking in environments influenced by collective feedback.

Additionally, our study did not reveal significant overconfidence among participants, a result that contrasts with survey-based studies which often report prevalent overconfidence in assessing information accuracy. In our experimental setup, participants' self-reported confidence closely matched their actual performance, indicating realistic self-assessment. This finding implies that feedback interventions aimed at reducing overconfidence may be less necessary in contexts with structured tasks similar to ours.

The experiment also revealed evidence of confirmation bias in verification preferences. Participants exhibited a tendency to pay for verification of information they initially believed to be true, indicating a selective preference for reinforcing rather than challenging initial beliefs. This selective verification suggests that interventions should focus on encouraging a balanced verification approach, promoting verification of both confirming and disconfirming information.

Moreover, the results indicate an asymmetry in utility perceptions. Participants showed greater concern for misclassifying false information as accurate than for the reverse, aligning with a utility framework that weights the costs of trusting misinformation more heavily. This asymmetry reflects real-world tendencies where the consequences of mistakenly accepting misinformation are perceived as more severe.

Finally, we found that political alignment influenced participants' classification judgments. Specifically, participants were more likely to judge headlines as inaccurate when they favored the government and conflicted with the participants' personal political views. This result underscores how ideological alignment shapes information processing and evaluation, with skepticism heightened for information that contradicts one's own ideological stance.

A. Implications and Future Research

Our findings suggest that feedback interventions, particularly those involving peer performance metrics, need careful calibration to avoid unintended decreases in verification demand. The lack of overconfidence observed in our experiment implies that individuals may already possess a reasonably accurate self-assessment of their information classification skills in structured settings. However, the observed confirmation bias in verification demand calls for further exploration into how individuals' motivations for verification might be nudged to encourage unbiased fact-checking.

In practice, developing effective strategies to counter misinformation may benefit from feedback mechanisms that balance accuracy with an emphasis on the importance of verification across all classifications, not just those aligning with preexisting beliefs. Future research could explore alternative feedback designs and assess the long-term impacts of feedback on verification behavior in more naturalistic settings, where stakes are higher, and feedback complexity can more closely reflect real-world challenges.

REFERENCES

Arin, K Peren, Deni Mazrekaj, and Marcel Thum, "Ability of detecting and willingness to share fake news," *Scientific Reports*, 2023, *7298.*

Arrow, K. J., D. Blackwell, and M. A. Girshick, "Bayes and Minimax Solutions of Sequential Decision Problems," *Econometrica*, 7 1949, *17*, 213.

Aslett, Kevin, Zeve Sanderson, William Godel, Nathaniel Persily, Jonathan Nagler, and Joshua A. Tucker, "Online searches to evaluate misinformation can increase its perceived veracity," *Nature*, 1 2024, *625*, 548–556.

Bateman, Jon and Dean Jackson, *Countering Disinformation Effectively: An Evidence-Based Policy Guide*, Carnegie Endowment for International Peace, 2024.

Charness, Gary, Nir Chemaya, and Dario Trujano-Ochoa, "Learning your own risk preferences," *Journal of Risk and Uncertainty*, 8 2023, *67*, 1–19.

_ , Uri Gneezy, and Vlastimil Rasocha, "Experimental methods: Eliciting beliefs," *Journal of Economic Behavior and Organization*, 2021, *189*, 234–256.

Danz, David, Lise Vesterlund, and Alistair J Wilson, "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, 2022, *112*, 2851–2883.

Eberlein, Marion, Sandra Ludwig, and Julia Nafziger, "The effects of feedback on self-assessment," *Bulletin of Economic Research*, 2011, *63*, 307–3378.

Erat, Serhat, Kurtulus Demirkol, and M Eyyüp Sallabas, "Overconfidence and its link with feedback," *Active Learning in Higher Education*, 2022, *3*, 173–187.

Ferraro, Paul J, "Know thyself: incompetence and overconfidence," *Experimental Laboratory Working Paper Series No. 2003-001. Department of Economics, Andrew Young School of Policy Studies, Georgia State University*, 2005.

Hoes, Emma, Brian Aitken, Jingwen Zhang, Tomasz Gackowski, and Magdalena Wojcieszak, "Prominent misinformation interventions reduce misperceptions but increase scepticism," *Nature Human Behavior*, 2023.

Kartal, Melis and Jean Robert Tyran, "Fake News, Voter Overconfidence, and the Quality of Democratic Choice," *American Economic Review*, 10 2022, *112*, 3367–97.

Kogelnik, Maria, "Performance Feedback and Gender Differences in Persistence," *SSRN Electronic Journal*, 12 2022.

Kozyreva, Anastasia, Philipp Lorenz-Spreen, Stefan M. Herzog, Ullrich K.H. Ecker, Stephan Lewandowsky, Ralph Hertwig, Ayesha Ali, Joe Bak-Coleman, Sarit Barzilai, Melisa Basol, Adam J. Berinsky, Cornelia Betsch, John Cook, Lisa K. Fazio, Michael Geers, Andrew M. Guess, Haifeng Huang, Horacio Larreguy, Rakoen Maertens, Folco Panizza, Gordon Pennycook, David G. Rand, Steve Rathje, Jason Reifler, Philipp Schmid, Mark Smith, Briony Swire-Thompson, Paula Szewach, Sander van der Linden, and Sam Wineburg, "Toolbox of individual-level interventions against online misinformation," *Nature Human Behaviour 2024 8:6*, 5 2024, *8*, 1044–1052.

Lyons, Benjamin A., Jacob M. Montgomery, Andrew M. Guess, Brendan Nyhan, and Jason Reifler, "Overconfidence in news judgments is associated with false news susceptibility," *Proceedings of the National Academy of Sciences of the United States of America*, 6 2021, *118.*

Moore, Don A. and Paul J. Healy, "The Trouble With Overconfidence," *Psychological Review*, 4 2008, *115*, 502–517.

Oprea, Ryan and Sevgi Yuksel, "Social Exchange of Motivated Beliefs," *Journal of the European Economic Association*, 2022, *20*, 667–699.

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in Political Behavior," *American Economic Review*, 2015, *105*, 504–535.

Pennycook, Gordon and David G. Rand, "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking," *Journal of Personality*, 4 2020, *88*, 185–200.

_ and _ , "Nudging Social Media toward Accuracy," *Annals of the American Academy of Political and Social Science*, 3 2022, *700*, 152–164.

_ , Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand, "Shifting attention to accuracy can reduce misinformation online," *Nature 592*, 2021, *592*, 590–595.

Pulford, Briony D. and Andrew M. Colman, "Overconfidence: Feedback and item difficulty effects," *Personality and Individual Differences*, 7 1997, *23*, 125–133.

Thaler, Michael, "The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News," *American Economic Journal: Microeconomics*, 2024, *16*, 1–38.

Vosoughi, Soroush, Deb Roy, and Sinan Aral, "The spread of true and false news online," *Science*, 2018, *359*, 1146–1151.

Wald, Abraham, "Foundations of a General Theory of Sequential Decision Functions," *Econometrica*, 10 1947, *15*, 279.

Wilson, Alistair J and Emanuel Vespa, "Paired-uniform scoring: Implementing a binarized scoring rule with non-mathematical language," 2018.

Mathematical Appendix

A1. *Expected Value of the Signal Considering Reclassification*

General Setup

1) **Initial Classification**: The agent initially classifies the headline as c (either accurate (a) or fake (f)).
2) **Receive Signal**: The agent receives a signal s which can either confirm or contradict their initial classification.
3) **Reclassification**: Based on the signal, the agent makes a new classification c'.

Expected Utility with Signal and Reclassification ($EU_{\text{SIGNAL}}^{\text{UPDATE}}(c)$)

The expected utility with the signal and reclassification is calculated by considering the updated posterior probabilities and the new classification based on the signal.

Step 1: Define Probabilities and Utilities

- **Initial Posterior Probabilities**:

$$P(A|c) = \frac{P(c|A) \cdot (1 - p_f)}{P(c)}, \quad P(F|c) = \frac{P(c|F) \cdot p_f}{P(c)}$$

where:

$$P(c) = (1 - p_f) \cdot P(c|A) + p_f \cdot P(c|F)$$

- **Signal Probabilities**:

$$q_a = P(s = a|A), \quad q_f = P(s = f|F)$$

- **Utilities**:

$$U_A, U_F, U_{AF}, U_{FA}$$

Step 2: Define Updated Posterior Probabilities Given Signal

After observing the signal, the agent updates their beliefs:

- **Posterior Probabilities Given Signal** $s = a$:

$$P(A|s = a, c) = \frac{q_a \cdot P(A|c)}{q_a \cdot P(A|c) + (1 - q_f) \cdot P(F|c)}$$

$$P(F|s = a, c) = \frac{(1 - q_f) \cdot P(F|c)}{q_a \cdot P(A|c) + (1 - q_f) \cdot P(F|c)}$$

- **Posterior Probabilities Given Signal** $s = f$:

$$P(A|s=f,c) = \frac{(1-q_a)\cdot P(A|c)}{(1-q_a)\cdot P(A|c) + q_f \cdot P(F|c)}$$

$$P(F|s=f,c) = \frac{q_f \cdot P(F|c)}{(1-q_a)\cdot P(A|c) + q_f \cdot P(F|c)}$$

Step 3: Calculate Expected Utility After Signal

We assume that the signal is strong enough (proposition 1) to assure that the best reclassification is to follow what the signal indicates is the state of the world. Otherwise, the signal would have no instrumental value, and therefore, $EVSI = 0$.

The expected utility with the signal, considering the possibility of reclassification, is:

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

Here, $P(s=a|c)$ and $P(s=f|c)$ are the probabilities of receiving the signals $s = a$ and $s = f$ given the initial classification c. These probabilities are determined by Bayes' rule, considering the agent's initial classification and the properties of the signal.

Given the initial classification c:

$$P(s=a|c) = q_a \cdot P(A|c) + (1-q_f)\cdot P(F|c)$$

$$P(s=f|c) = (1-q_a)\cdot P(A|c) + q_f \cdot P(F|c)$$

Expected Utility After Signal $s = a$

$$EU_{\text{new classification}}(s=a,c) = P(A|s=a,c)\cdot U_A + P(F|s=a,c)\cdot U_{AF}$$

Expected Utility After Signal $s = f$

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

Step 4: Combine Expected Utilities

$$EU^{\text{update}}_{\text{signal}}(c) = [q_a \cdot P(A|c) + (1-q_f) \cdot P(F|c)] \cdot [P(A|s=a,c) \cdot U_A + P(F|s=a,c) \cdot U_{AF}]$$

$$+ [(1-q_a) \cdot P(A|c) + q_f \cdot P(F|c)] \cdot [P(F|s=f,c) \cdot U_F + P(A|s=f,c) \cdot U_{FA}]$$

Step 5: Expected Value of the Signal (EVSI)

Finally, the EVSI is the difference between the expected utility with the signal and the expected utility without the signal.

$$EVSI = EU^{\text{update}}_{\text{signal}}(c) - EU_{\text{no signal}}(c)$$

Conclusion

By allowing the agent to update their classification based on the signal, we account for the dynamic decision-making process. The expected value of the signal (EVSI) is derived by comparing the expected utility with the signal (considering reclassification) to the expected utility without the signal. This approach shows the impact of additional information on improving decision-making accuracy.

A2. Proof of Proposition 2: Need for a Strong Enough Signal

To ensure that people follow the signal S for reclassification, we must prove that the expected utility of reclassifying based on the signal is higher than not reclassifying. Without loss of generality, we will first consider the case where the initial classification $c = f$ and the signal $s = a$.

Initial Setup

1) **Initial Classification**: $c = f$ (classified as fake)
2) **Signal Received**: $s = a$ (signal indicates accurate)

We need to show that reclassifying the headline as accurate ($c' = a$) based on the signal is optimal.

Expected Utility of Not Reclassifying

If the agent does not reclassify and sticks with the initial classification $c = f$, but knows the signal $s = a$, the expected utility is:

$$EU_{\text{no reclassification}}(f, s=a) = P(A|s=a,f) \cdot U_{FA} + P(F|s=a,f) \cdot U_F$$

Expected Utility of Reclassifying

If the agent reclassifies the headline based on the signal $s = a$, the expected utility is:

$$EU_{\text{reclassification}}(f, s = a) = P(A|s = a, f) \cdot U_A + P(F|s = a, f) \cdot U_{AF}$$

Posterior Probabilities

The posterior probabilities given the signal $s = a$ and initial classification $c = f$ are:

$$P(A|s = a, f) = \frac{q_a \cdot P(A|f)}{q_a \cdot P(A|f) + (1 - q_f) \cdot P(F|f)}$$
$$P(F|s = a, f) = \frac{(1 - q_f) \cdot P(F|f)}{q_a \cdot P(A|f) + (1 - q_f) \cdot P(F|f)}$$

Condition for Reclassifying

To prove that reclassifying based on the signal is optimal, we need:

$$EU_{\text{reclassification}}(f, s = a) > EU_{\text{no reclassification}}(f, s = a)$$

Substituting the utilities, we get:

$$P(A|s = a, f) \cdot U_A + P(F|s = a, f) \cdot U_{AF} > P(A|s = a, f) \cdot U_{FA} + P(F|s = a, f) \cdot U_F$$

Given the simplifying assumptions:

$$U_A = 1, \quad U_F = 1, \quad U_{AF} = 0, \quad U_{FA} = 0$$

The inequality simplifies to:

$$P(A|s = a, f) \cdot 1 + P(F|s = a, f) \cdot 0 > P(A|s = a, f) \cdot 0 + P(F|s = a, f) \cdot 1$$

This reduces to:

$$P(A|s = a, f) > P(F|s = a, f)$$

Verifying the Posterior Probabilities

Substitute the posterior probabilities:

$$\frac{q_a \cdot P(A|f)}{q_a \cdot P(A|f) + (1-q_f) \cdot P(F|f)} > \frac{(1-q_f) \cdot P(F|f)}{q_a \cdot P(A|f) + (1-q_f) \cdot P(F|f)}$$

Since the denominators are the same, we can simplify this to:

$$q_a \cdot P(A|f) > (1-q_f) \cdot P(F|f)$$

Since $P(F|f) = 1 - P(A|f)$, we have:

$$q_a \cdot P(A|f) > (1-q_f) \cdot (1 - P(A|f))$$

Expanding and rearranging terms, we get:

$$q_a \cdot P(A|f) > (1-q_f) - (1-q_f) \cdot P(A|f)$$

$$q_a \cdot P(A|f) + (1-q_f) \cdot P(A|f) > (1-q_f)$$

$$P(A|f) \cdot (q_a + 1 - q_f) > (1-q_f)$$

Dividing both sides by $(q_a + 1 - q_f)$:

$$P(A|f) > \frac{1-q_f}{q_a + 1 - q_f}$$

This shows that the signal needs to be strong enough such that q_a is sufficiently large compared to $1 - q_f$, ensuring that the agent reclassifies the headline as accurate based on the signal. This proves that a strong signal is necessary to ensure that people follow the signal S for reclassification.

Trivial Case: $c = s = a$

If the initial classification $c = a$ and the signal $s = a$, then reclassification is not necessary because the initial classification is already accurate. The expected utility remains the same:

$$EU_{\text{reclassification}}(a, s=a) = P(A|s=a, a) \cdot U_A + P(F|s=a, a) \cdot U_{AF}$$

Given the simplifying assumptions, this reduces to:

$$EU_{\text{reclassification}}(a, s=a) = P(A|s=a, a) \cdot 1 + P(F|s=a, a) \cdot 0 = P(A|s=a, a)$$

The expected utility of not reclassifying is:

$$EU_{\text{no reclassification}}(a, s=a) = P(A|s=a, a) \cdot U_A + P(F|s=a, a) \cdot U_{AF}$$

Given the simplifying assumptions, this reduces to:

$$EU_{\text{no reclassification}}(a, s = a) = P(A|s = a, a) \cdot 1 + P(F|s = a, a) \cdot 0 = P(A|s = a, a)$$

Since both expected utilities are equal, reclassification is trivial in this case.

Other Cases

The same process follows for the cases $c = a, s = f$ and $c = s = f$. For these cases, the conditions are as follows:

1) **Case** $c = a, s = f$:

$$EU_{\text{reclassification}}(a, s = f) > EU_{\text{no reclassification}}(a, s = f)$$

Substituting the utilities, we get:

$$P(F|s = f, a) \cdot U_F + P(A|s = f, a) \cdot U_{FA} > P(F|s = f, a) \cdot U_{AF} + P(A|s = f, a) \cdot U_A$$

Given the simplifying assumptions:

$$P(F|s = f, a) \cdot 1 + P(A|s = f, a) \cdot 0 > P(F|s = f, a) \cdot 0 + P(A|s = f, a) \cdot 1$$

This reduces to:

$$P(F|s = f, a) > P(A|s = f, a)$$

Verifying the posterior probabilities:

$$\frac{q_f \cdot P(F|a)}{q_f \cdot P(F|a) + (1 - q_a) \cdot P(A|a)} > \frac{(1 - q_a) \cdot P(A|a)}{q_f \cdot P(F|a) + (1 - q_a) \cdot P(A|a)}$$

Since the denominators are the same, we can simplify this to:

$$q_f \cdot P(F|a) > (1 - q_a) \cdot P(A|a)$$

Since $P(A|a) = 1 - P(F|a)$, we have:

$$q_f \cdot P(F|a) > (1 - q_a) \cdot (1 - P(F|a))$$

Expanding and rearranging terms, we get:

$$q_f \cdot P(F|a) > (1 - q_a) - (1 - q_a) \cdot P(F|a)$$

$$q_f \cdot P(F|a) + (1 - q_a) \cdot P(F|a) > (1 - q_a)$$
$$P(F|a) \cdot (q_f + 1 - q_a) > (1 - q_a)$$

Dividing both sides by $(q_f + 1 - q_a)$:

$$P(F|a) > \frac{1 - q_a}{q_f + 1 - q_a}$$

2) **Case** $c = s = f$: If the initial classification $c = f$ and the signal $s = f$, then reclassification is unnecessary because the initial classification is already correct. The expected utility remains the same:

$$EU_{\text{reclassification}}(f, s = f) = P(F|s = f, f) \cdot U_F + P(A|s = f, f) \cdot U_{AF}$$

Given the simplifying assumptions, this reduces to:

$$EU_{\text{reclassification}}(f, s = f) = P(F|s = f, f) \cdot 1 + P(A|s = f, f) \cdot 0 = P(F|s = f, f)$$

The expected utility of not reclassifying is:

$$EU_{\text{no reclassification}}(f, s = f) = P(F|s = f, f) \cdot U_F + P(A|s = f, f) \cdot U_{AF}$$

Given the simplifying assumptions, this reduces to:

$$EU_{\text{no reclassification}}(f, s = f) = P(F|s = f, f) \cdot 1 + P(A|s = f, f) \cdot 0 = P(F|s = f, f)$$

Since both expected utilities are equal, reclassification is trivial in this case.

Conditions

Therefore, to ensure that the signal is strong enough to prompt optimal reclassification in both cases, we need to satisfy two key conditions:

$$P(A|f) > \frac{1 - q_f}{q_a + 1 - q_f}$$
$$P(A|a) < \frac{q_f}{q_f + 1 - q_a}$$

Considering the conditions for proposition 1 we have that,

$$\frac{1 - q_f}{q_a + 1 - q_f} < P(A|f) < P(A|a) < \frac{q_f}{q_f + 1 - q_a}$$

A3. Proof of Proposition 1: Sufficient and Necessary Condition for $P(A|c = f) < P(A) < P(A|c = a)$ and $P(F|c = a) < P(F) < P(F|c = f)$

Proof of Necessity and Sufficiency

We will prove that $1 < \frac{P(c=a|A)}{P(c=a|F)}$ and $1 < \frac{P(c=f|F)}{P(c=f|A)}$ if and only if $P(A|c = f) < P(A) < P(A|c = a)$ and $P(F|c = a) < P(F) < P(F|c = f)$.

B1. Definitions and Setup

Let:

- $P(A)$ be the prior probability that the state is accurate.
- $P(F)$ be the prior probability that the state is fake.
- $P(c = a|A)$ be the probability of classifying a headline as accurate given it is accurate.
- $P(c = a|F)$ be the probability of classifying a headline as accurate given it is fake.
- $P(c = f|F)$ be the probability of classifying a headline as fake given it is fake.
- $P(c = f|A)$ be the probability of classifying a headline as fake, given it is accurate.

B2. Posterior Probabilities

The posterior probabilities after observing the classification c are given by:

- Posterior probability of A given $c = f$:

$$P(A|c = f) = \frac{P(c = f|A) \cdot P(A)}{P(c = f|A) \cdot P(A) + P(c = f|F) \cdot P(F)}$$

- Posterior probability of A given $c = a$:

$$P(A|c = a) = \frac{P(c = a|A) \cdot P(A)}{P(c = a|A) \cdot P(A) + P(c = a|F) \cdot P(F)}$$

- Posterior probability of F given $c = f$:

$$P(F|c = f) = \frac{P(c = f|F) \cdot P(F)}{P(c = f|A) \cdot P(A) + P(c = f|F) \cdot P(F)}$$

- Posterior probability of F given $c = a$:

$$P(F|c = a) = \frac{P(c = a|F) \cdot P(F)}{P(c = a|A) \cdot P(A) + P(c = a|F) \cdot P(F)}$$

B3. Part 1: Sufficiency ($\Rightarrow$)

Assume that $1 < \frac{P(c=a|A)}{P(c=a|F)}$ and $1 < \frac{P(c=f|F)}{P(c=f|A)}$. We want to show that this implies $P(A|c = f) < P(A) < P(A|c = a)$ and $P(F|c = a) < P(F) < P(F|c = f)$.

Analyze the Posterior Probabilities

1) **For $P(A|c=f)$:**

Given the condition $1 < \frac{P(c=f|F)}{P(c=f|A)}$, we know that:

$$\frac{P(c=f|F)}{P(c=f|A)} > 1$$

This implies $P(c=f|F) > P(c=f|A)$. As a result, in the posterior probability expression:

$$P(A|c=f) = \frac{P(c=f|A)\cdot P(A)}{P(c=f|A)\cdot P(A) + P(c=f|F)\cdot P(F)}$$

The denominator $P(c=f|A)\cdot P(A)+P(c=f|F)\cdot P(F)$ will be larger than the numerator $P(c=f|A)\cdot P(A)$, causing $P(A|c=f)$ to be smaller than the prior $P(A)$. Therefore:

$$P(A|c=f) < P(A)$$

2) **For $P(A|c=a)$:**

Given the condition $1 < \frac{P(c=a|A)}{P(c=a|F)}$, we know that:

$$\frac{P(c=a|A)}{P(c=a|F)} > 1$$

This implies $P(c=a|A) > P(c=a|F)$. As a result, in the posterior probability expression:

$$P(A|c=a) = \frac{P(c=a|A)\cdot P(A)}{P(c=a|A)\cdot P(A) + P(c=a|F)\cdot P(F)}$$

The numerator $P(c=a|A)\cdot P(A)$ will dominate the denominator $P(c=a|A)\cdot P(A)+P(c=a|F)\cdot P(F)$, causing $P(A|c=a)$ to be larger than the prior $P(A)$. Therefore:

$$P(A|c=a) > P(A)$$

3) **For $P(F|c=f)$ and $P(F|c=a)$:**

Similarly, the same reasoning applies to $P(F|c=f)$ and $P(F|c=a)$, given that:

$$\frac{P(c=f|F)}{P(c=f|A)} > 1 \quad \text{and} \quad \frac{P(c=a|A)}{P(c=a|F)} > 1$$

This implies that:

$$P(F|c = a) < P(F) < P(F|c = f)$$

B4. Part 2: Necessity ($\Leftarrow$)

Assume that $P(A|c = f) < P(A) < P(A|c = a)$ and $P(F|c = a) < P(F) < P(F|c = f)$. We need to show that this implies $1 < \frac{P(c=a|A)}{P(c=a|F)}$ and $1 < \frac{P(c=f|F)}{P(c=f|A)}$.

Analyzing the Posterior Probabilities

- **For $P(A|c = f) < P(A)$:**

  Given the posterior probability expression:

$$P(A|c = f) = \frac{P(c = f|A) \cdot P(A)}{P(c = f|A) \cdot P(A) + P(c = f|F) \cdot P(F)}$$

  If $P(A|c = f) < P(A)$, then the likelihood ratio $\frac{P(c=f|F)}{P(c=f|A)}$ must be greater than 1. This is because the posterior $P(A|c = f)$ being less than $P(A)$ implies that the signal $c = f$ is more likely to come from the fake state F, meaning:

$$\frac{P(c = f|F)}{P(c = f|A)} > 1$$

- **For $P(A|c = a) > P(A)$:**

  Given the posterior probability expression:

$$P(A|c = a) = \frac{P(c = a|A) \cdot P(A)}{P(c = a|A) \cdot P(A) + P(c = a|F) \cdot P(F)}$$

  If $P(A|c = a) > P(A)$, then the likelihood ratio $\frac{P(c=a|A)}{P(c=a|F)}$ must be greater than 1. This is because the posterior $P(A|c = a)$ being greater than $P(A)$ implies that the signal $c = a$ is more likely to come from the accurate state A, meaning:

$$\frac{P(c = a|A)}{P(c = a|F)} > 1$$

- **For $P(F|c = f) > P(F)$ and $P(F|c = a) < P(F)$:**

  By symmetry, the same reasoning applies for $P(F|c = f) > P(F)$ and $P(F|c = a) < P(F)$. The likelihood ratios $\frac{P(c=f|F)}{P(c=f|A)} > 1$ and $\frac{P(c=a|A)}{P(c=a|F)} > 1$ are necessary conditions to satisfy these posterior inequalities.

B5. Conclusion

Thus, we have shown that: $1 < \frac{P(c=a|A)}{P(c=a|F)}$ and $1 < \frac{P(c=f|F)}{P(c=f|A)}$ are necessary and sufficient conditions for:

$$P(A|c = f) < P(A) < P(A|c = a)$$

$$P(F|c = a) < P(F) < P(F|c = f)$$

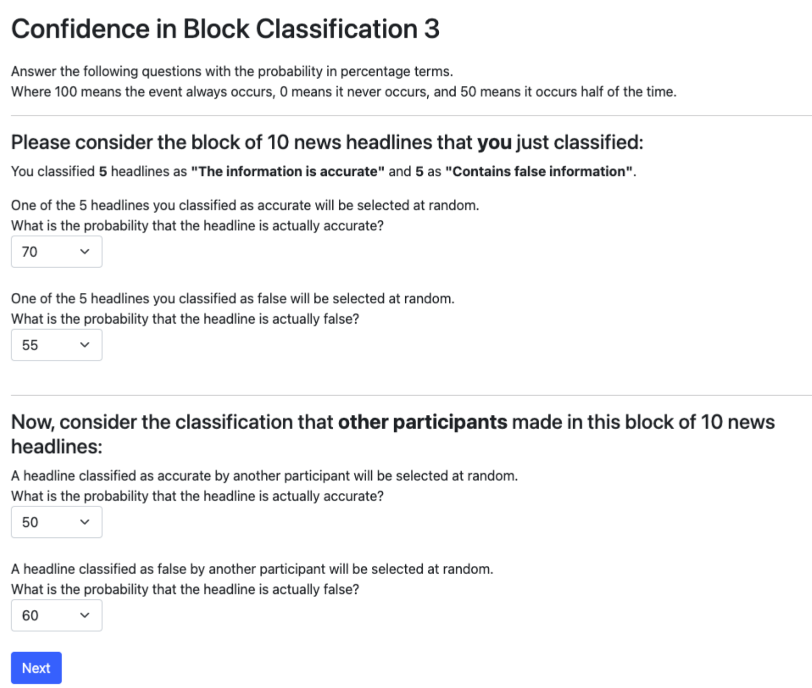
Corollary: Informativeness of the Signal

The same analysis can be applied to the signal. Therefore $1 < \frac{P(s=a|A)}{P(s=a|F)}$ and $1 < \frac{P(s=f|F)}{P(s=f|A)} \iff$

$$P(A|s = f) < P(A) < P(A|s = a)$$
$$P(F|s = a) < P(F) < P(F|s = f)$$

Materials

C1. Confidence Elicitation


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 the 5 headlines you classified as accurate will be selected at random.
What is the probability that the headline is actually accurate?

70

One of the 5 headlines you classified as false will be selected at random.
What is the probability that the headline is actually false?

55

Now, consider the classification that **other participants** made in this block of 10 news headlines:

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

60

Next

Figure C1. Screenshot of the translated Confidence elicitation as seen by the participants.

C2. Headlines Used in the Experiment

Block	Real	Headline	Translated Headline
1	1	Se inaugura un nuevo museo en honor a Cantinflas en la Ciudad de México	A New Museum in Honor of Cantinflas is Inaugurated in Mexico City
1	1	El salario mínimo en México se incrementa 20% en 2024	Minimum Wage in Mexico Increases by 20% in 2024
1	1	La variante Ómicron es la única de preocupación que circula a nivel mundial; es más transmisible, aunque menos peligrosa que la variante Delta	The Omicron Variant is the Only Variant of Concern Circulating Worldwide; It Is More Transmissible, Though Less Dangerous than the Delta Variant
1	1	Se suspende programa humanitario para trabajar o solicitar asilo en Estados Unidos para Haití, Venezuela, Nicaragua y Cuba	Humanitarian program to work or apply for asylum in the United States for Haiti, Venezuela, Nicaragua, and Cuba is suspended
1	1	Incrementó en el uso de energías renovables en México	Increase in the Use of Renewable Energy in Mexico
1	0	Turismo internacional se desploma en 2024, México ya no es un destino atractivo	International Tourism Collapses, Mexico is No Longer an Attractive Destination
1	0	Luto en México por accidente aéreo de un avión de pasajeros. No hubo sobrevivientes	National Mourning in Mexico. Terrible Passenger Plane Crash in 2024. No Survivors
1	1	En 2024, se intensificaron los incendios forestales en México	In 2024, Wildfires Intensified in Various Regions of Mexico
1	1	Existen programas de apoyo a pequeñas empresas lanzados por el gobierno mexicano	There Are Support Programs for Small Businesses Launched by the Mexican Government
1	0	Se firma un tratado del Foro Económico Mundial que busca reconocer la pedofilia como orientación sexual	A World Economic Forum Treaty Is Signed to Recognize Pedophilia as a Sexual Orientation
2	1	El INAH cobrará $60 por tomar fotografías para uso comercial en museos y sitios arqueológicos	INAH Will Charge $60 for Taking Photos in Museums and Archaeological Sites for Commercial Use
2	0	Se publica la lista de apellidos que pueden solicitar la ciudadanía española	Spain Publishes a List of Surnames That Allow One to Apply for Spanish Citizenship

2	0	En Irán censuraron los Juegos Olímpicos; todas las mujeres aparecen con rectángulos o asteriscos cubriéndolas	Iran Censored the Olympics; All Women Appear with Rectangles or Asterisks Covering Them
2	0	Iniciará juicio en contra de la ministra presidenta de la SCJN por participar en el paro de trabajadores del Poder Judicial.	Trial Against the Chief Justice of the Supreme Court for Participating in the Judicial Workers' Strike to Begin
2	0	La Organización de Estados Americanos (OEA) sanciona a México por dar asilo a Jorge Glass en la embajada mexicana en Ecuador	The Organization of American States (OAS) Managed to Sanction Mexico for Granting Asylum to Jorge Glass in the Mexican Embassy
2	0	El hijo de Nicolás Maduro es captado en video manejando un Ferrari dorado	Nicolás Maduro's son is seen driving a golden Ferrari
2	1	El Ozempic, promovido en redes para bajar de peso, es un tratamiento controlado para la diabetes tipo 2	Ozempic Is Actually a Controlled Treatment for Type 2 Diabetes
2	1	México alcanza cifra récord en exportaciones agrícolas	Mexico Reaches Record High in Agricultural Exports
2	0	Atletas ucranianos portaron pulseras de tobillo con GPS para evitar su huida después de los Juegos Olímpicos de París 2024	Ukrainian Athletes Wore GPS Ankle Bracelets to Prevent Them from Fleeing After the Olympic Games
2	1	Ningún país ha declarado confinamiento por mpox tras la nueva emergencia sanitaria anunciada por la OMS	No country has declared a lockdown due to mpox following the new health emergency announced by the WHO
3	0	La cama "antisexo" siguió siendo utilizada en los Juegos Olímpicos de París 2024	The "Anti-Sex" Bed Will Continue to Be Used at the Paris 2024 Olympics
3	0	El aeropuerto de Suecia fue descontaminado debido a contagios de Mpox	Sweden's Airport Was Decontaminated Due to Mpox Infections
3	0	Miss Venezuela protestó ante las cámaras contra su gobierno en una alfombra roja	Miss Venezuela Protested Against Her Government on a Red Carpet
3	1	Avances en la investigación de nuevas vacunas desarrolladas en México	Advances in the Research of New Vaccines Developed in Mexico

3	0	Un estadounidense se suicidó saltando desde su habitación durante el Baja Beach Fest 2024 en México	An American Committed Suicide During the Baja Beach Fest 2024 in Mexico
3	1	Descubrimiento de nuevas ruinas mayas en la península de Yucatán en 2024	Discovery of New Mayan Ruins in the Yucatán Peninsula
3	1	México termina en el puesto 65 del medallero en los Juegos Olímpicos de París 2024	Mexico Finishes 65th in the Medal Table at the Paris 2024 Olympic Games
3	1	Mexico envió dos aviones en 2023 para rescatar connacionales varados en Israel por el conflicto en Gaza	Sedena and SRE Sent Two Planes to Rescue Mexicans Stranded in Israel
3	0	Consumir alimentos alcalinos ayuda a contrarrestar la variante Omicron del coronavirus	Maintaining a pH (Acidity Level) Above 5.5 Can Prevent Covid-19 Infection
3	0	El Consejo para Prevenir y Eliminar la Discriminación (COPRED) busca suspender la celebración del Día del Padre en los centros educativos	COPRED Urges Elementary Schools Not to Exclude Children from Non-Normative Families on Father's Day
4	1	México tiene la tasa más baja de desempleo de la OCDE	Mexico Has the Lowest Unemployment Rate in the OECD
4	0	Gobierno de México entrega el nuevo "Bono Mujeres" por 2 mil 700 pesos	Mexican Government Issues the New "Women's Bonus" for 2,700 Pesos
4	1	Se registró una disminución de 5.1 millones personas en pobreza en el actual gobierno	A Decrease of 5.1 Million People in Poverty Was Recorded During the Current Government
4	0	Tribunal Electoral encuentra irregularidades graves en el triunfo de Claudia Sheinbaum	Electoral Tribunal Finds Serious Irregularities in Claudia Sheinbaum's Victory
4	1	La presidenta electa Claudia Sheinbaum anuncia beca universal para estudiantes de nivel básico	President-Elect Claudia Sheinbaum Announces Universal Scholarship for Elementary School Students
4	0	El Tren Maya se completará sin impacto ambiental, según estudios científicos independientes	The Maya Train Will Be Completed Without Environmental Impact, According to Independent Scientific Studies
4	1	11 Ministros de la Suprema Corte de Justicia de la Nación ganan $206,246 pesos mensuales netos	11 Supreme Court Justices Earn $206,246 Pesos Monthly Net

4	1	México envió dos aviones a Israel para rescatar a las y los mexicanos varados por el conflicto con Palestina.	Mexico Sent Two Planes to Israel to Rescue Mexicans Stranded Due to the Conflict with Palestine
4	1	En solo ocho de cada 100 delitos en México se abre una carpeta de investigación	Only Eight Out of Every 100 Crimes in Mexico Lead to an Investigation Being Opened
4	0	México prepara una reunión con los presidentes de Rusia y Corea del Norte para comprar armas	Mexico Prepares a Meeting with the Presidents of Russia and North Korea to Buy Weapons
5	1	En el gobierno de AMLO se logró una reducción en la tasa de homicidios.	AMLO's Government Achieved a Reduction in the Homicide Rate
5	1	Primer director femenino de la CFE es nombrado en México	First Female Director of the CFE Is Appointed in Mexico
5	0	Durante el gobierno de Andrés Manuel López Obrador, la deuda pública subió 64%	During Andrés Manuel López Obrador's Government, Public Debt Increased by 64%
5	0	Poder Judicial de la Federación hay 53,737 personas que ganan más que el Presidente	In the Federal Judiciary, 53,737 People Earn More Than the President
5	1	La pobreza extrema en México incrementó de 2018 a 2022	Extreme Poverty in Mexico Increased from 2018 to 2022
5	1	Hay déficit presupuestario en el 2024 por parte del gobierno de México	Mexico's Government Faces a Budget Deficit in 2024
5	1	EU critica al Gobierno de AMLO por 'desacreditar a periodistas'	U.S. Criticizes AMLO's Government for 'Discrediting Journalists'
5	1	Sedena gastó más que el presupuesto autorizado por el Congreso	Sedena Spent More Than the Budget Authorized by Congress
5	0	México ya produce el 90% de la gasolina que consume, como afirma AMLO	Mexico Now Produces 90% of the Gasoline It Consumes, As Claimed by AMLO
5	0	Metro de CDMX dejará de ser gratis para adultos mayores	CDMX Metro Will No Longer Be Free for Senior Citizens