# "TRUE" labels on Fact-Checking interventions increases sharing of corrections on vaccine misinformation

Natalia Aruguete* Flavia Batista † Ernesto Calvo ‡ Carlos Scartascini §
Tiago Ventura ¶

August 28, 2022

**Abstract**

There is significant research that explains the users' decision to share misinformation online. However, we know less about the reasons that explain the decision to share fact checks. In this article we analyze the effect of *confirmation* and *refutation* frames on the decision to *like*, *share*, and *comment* fact checks online. We randomly expose respondents to semantically equivalent content that is worded as a confirmation of accurate information ("it is TRUE that vaccines prevent against COVID-19") or as a refutation of misinformation ("it is FALSE that vaccines DO NOT prevent against COVID-19"). Respondents like, share, and reply *confirmation* frames at higher rates than *refutation* frames, even if the statements are semantically equivalent. Findings are important in designing policy interventions that optimize fact checking exposure, either increasing the amplification of the factually correct content or by reducing the salience of the issue when it is socially harmful. This is particularly relevant when addressing misinformation on topics such as health and toxic speech

*Universidad Nacional de Quilmes, UNQ. Email: nataliaaruguete@gmail.com. Webpage: http://unq.academia.edu/nataliaaruguete

†University of Maryland, Government and Politics, UMD. Address: 4118 Chiconteague, College Park, MD 20742, USA. Email: fbatista@umd.edu.

‡University of Maryland, Government and Politics, UMD. Address: 3140 Tydings Hall, College Park, MD 20742, USA. Email: ecalvo@umd.edu. Webpage: http://gvptsites.umd.edu/calvo/

§IADB. 1300 New York Avenue, N.W., Washington, DC 20577, USA. CARLOSSC@iadb.org.

¶University of Maryland, Government and Politics, UMD. Address: 4118 Chiconteague, College Park, MD 20742, USA. Email: venturat@umd.edu. Webpage: http://tiagoventura.rbind.io/

Fact-checking is today the first line of defence against misinformation. It is frequently defined as "the practice of systematically publishing assessments of the validity of claims made by public officials and institutions with an explicit attempt to identify whether a claim is factual" (Walter et al., 2020, p. 350). Research shows that fact-checks successfully influences people's discernment of misinformation claims and update their beliefs upon being corrected, both in survey and field experiments, and across different countries (Arechar et al., 2022; Porter and Wood, 2021; Bode and Vraga, 2015; Clayton et al., 2020).

In curving the spread of misinformation, fact-checkers may choose to pursue two very different strategies: they may publish *confirmation frames* that replace misinformation with the correct information or, alternatively, they may publish *refutation frames* using warning labels that tag content as misinformation. Selecting the first alternative amplifies factually *true* content while the second alternative warns about sharing *false* content.

Because fact-checks ability to curve the spread of misinformation depends on reaching its target audience, understanding which corrective content is more broadly accepted and shared by social media users is central to the fact checker's mission. The visibility of fact-checking messages rests on "the extent to which individuals share primarily attitude-consistent content with their social networks" (Shin and Thorson, 2017, p.).

The use of TRUE or FALSE frames by fact-checkers can be an important moderator of sharing that has received little attention in the past (Aruguete et al., 2022). There is a wealth of studies measuring the spread of factually *true* or *false* content in social media (Bode and Vraga, 2015; Del Vicario et al., 2016; Van Der Linden et al., 2017; Lazer et al., 2018), but little research that measures the spread of fact checks framed as TRUE or FALSE. The lack of research measuring the effect of *confirmation (TRUE)* and *refutation (FALSE)* frames is

surprising, because labeling content is the most important framing decision made by fact checking organizations. This important framing effect is distinct from other moderators such as cognitive congruence, affective activation, and partisanship (Grinberg et al., 2019; Lazer et al., 2018).

The choice of TRUE or FALSE frames is central to the activity of the fact-checker. As noted by Shin and Thorson (2017), "[u]nlike traditional journalism, which emphasizes detached objectivity and adheres to the 'he said, she said' style of reporting, contemporary fact-checking directly engages in adjudicating factual disputes by publicly deciding whose claim is correct or incorrect" (Shin and Thorson, 2017, p.1). How to decide a factual dispute, using *confirmation* or *refutation* frames, is an editorial choice that is independent of the source material that is being corrected (Vosoughi et al., 2018). Still, fact checkers more frequently use refutation frames (e.g. "It is FALSE"). The Washington Post's "pinnochios" are a good example of an editorial decision to grade messages according to how false they are. The heavy leaning in favor of refutation frames follows from a well established journalistic tradition of publishing *erratums*, corrections that are similar in length but contrary to original content that was factually incorrect. However, "[p]ost-publication retractions and corrections often fail to eliminate the influence of misinformation. In some cases, they reinforce falsehoods simply by repeating them"(Van Der Linden et al., 2017, 1141). While information can be accurate or inaccurate, the decision to frame a fact check as a confirmation (TRUE) or as a refutation (FALSE) is an editorial choice.

While identical in appearance, the confirmation frame "It is TRUE that vaccines prevent against COVID-19" results in a different behavioral response than the more frequently used refutation frame "It is FALSE that vaccines DO NOT prevent against COVID-19". The frame "It is TRUE that vaccines prevent" is more widely liked and shared. Respondents also report more positive emotions such as joy and optimism. By contrast, the frame "It is FALSE that

vaccines DO NOT prevent" is less liked and shared, and correlates with negative emotions such as anger and disgust. Although positive frames yield higher amplification rates than negative ones, refutation frames such as "It is FALSE that" are the most used frame by fact checkers around the globe.

Our experiment exposes a national representative sample of Argentine respondents to a Facebook post framed as a confirmation of accurate information ("it is TRUE that the Moderna vaccine prevents against the new Omicron variant") or as a refutation of misinformation ("it is FALSE that the Moderna vaccine does not prevent against the new Omicron variant"). The experiment rotates three different vaccines (Moderna, AstraZeneca, and Sputnik V), and the confirmation (TRUE) or refutation (FALSE) frames.

Our main hypotheses thus test whether and how TRUE (confirmation) or FALSE (refutation) interventions influence the on-line spread of misinformation corrections. Following from Aruguete et al. (2022), we expect respondents to share confirmation frames (TRUE) at higher rates than refutation frames (FALSE) [hypothesis 1 (H1)]. This effect is independent of other pro-attitudinal and counter-attitudinal preferences for sharing a correction, such as the effect of cognitive congruence, affective activation, and partisan attachment.

We expect negative frames to reduce the intent to share via two different mechanisms, which we test for explicitly. First, it has been shown that negation carries a heavier cognitive burden (Christensen, 2020). A significant literature in cognitive linguistics and cognitive psychology has previously documented differences in how we process positive or negative propositions that are semantically equivalent. Kaup et al. (2006) shows that individuals are faster to process statement such as "the umbrella was open" compared to its semantically equivalent "the umbrella was not closed". Subjects also display faster response times for "the umbrella was closed"

compared to "the umbrella was not open", because cognitive effort is not the result of the state of the umbrella (i.e. *open* or *closed*) but the result of how we process negation statements. In social media, a higher cognitive burden should prevent a fast, automatic, and affective response (Kahneman, 2011; Aruguete and Calvo, 2018), and ensure more evaluative sharing behavior consistent with lower amplification. We expect refutation frames to impose higher cognitive burden for respondents, therefore, associated with longer reading time [hypothesis 2 (H2)].

Second, we expect that confirmation of pro-attitudinal beliefs will carry a higher positive valence compared to the refutation of a counter-attitudinal belief. Confirmation statements such as "it is TRUE" convey to users the idea that this content is socially accepted and, thereby, less likely to expose them to public scrutiny and criticism. Therefore, confirmation frames have higher social validation, emphasizing a royal "WE" that collectively agrees with the proposition. In contrast, refutation frames such as "It is FALSE" indicate a lack of consensus and the potential for conflict. Refutation frames hint that there are at least some individuals or groups with competing beliefs. Further, the refutation frames places the emphasis of the correction on the offending group and their intent to spread misinformation. Therefore, the refutation of a counter-attitudinal belief is socially disputed, with an emphasis on the aggravating out-group with which "we", the in-group, do not agree. Thus, we hypothesize that confirmation frames will generate positive emotional reactions, and refutations will increases negative affective responses to the correction [hypothesis 3 (H3)].

The statement "it is TRUE that" is expected to increase sharing behavior compared to the statement "it is FALSE that" because the former is both cognitively easier to process and because sharing the TRUE message is socially acceptable. By contrast, "it is FALSE" is cognitively difficult and sharing this message aligns us with an in-group that is in conflict with an out-

group.

**From Theory to Design**

The two-arm design exposes respondents to a Facebook post that randomly confirms a clinically correct statement or refutes a clinically incorrect statement.[1] We consider three different vaccines (Sputnik V, Moderna, and AstraZeneca), to test if differences in the perceived quality of the vaccine have an effect. Respondents are randomly assigned to one of six treatment groups (i.e. confirmation or refutation frames for Moderna, AstraZeneca, or Sputnik V).

[1]It is important to highlight that the experiment is not providing misinformation to the respondents. Both confirmations and refutations are propositions that report on the efficacy of the vaccines against the variant Omicron. Therefore, respondents were not treated to misinformation content.

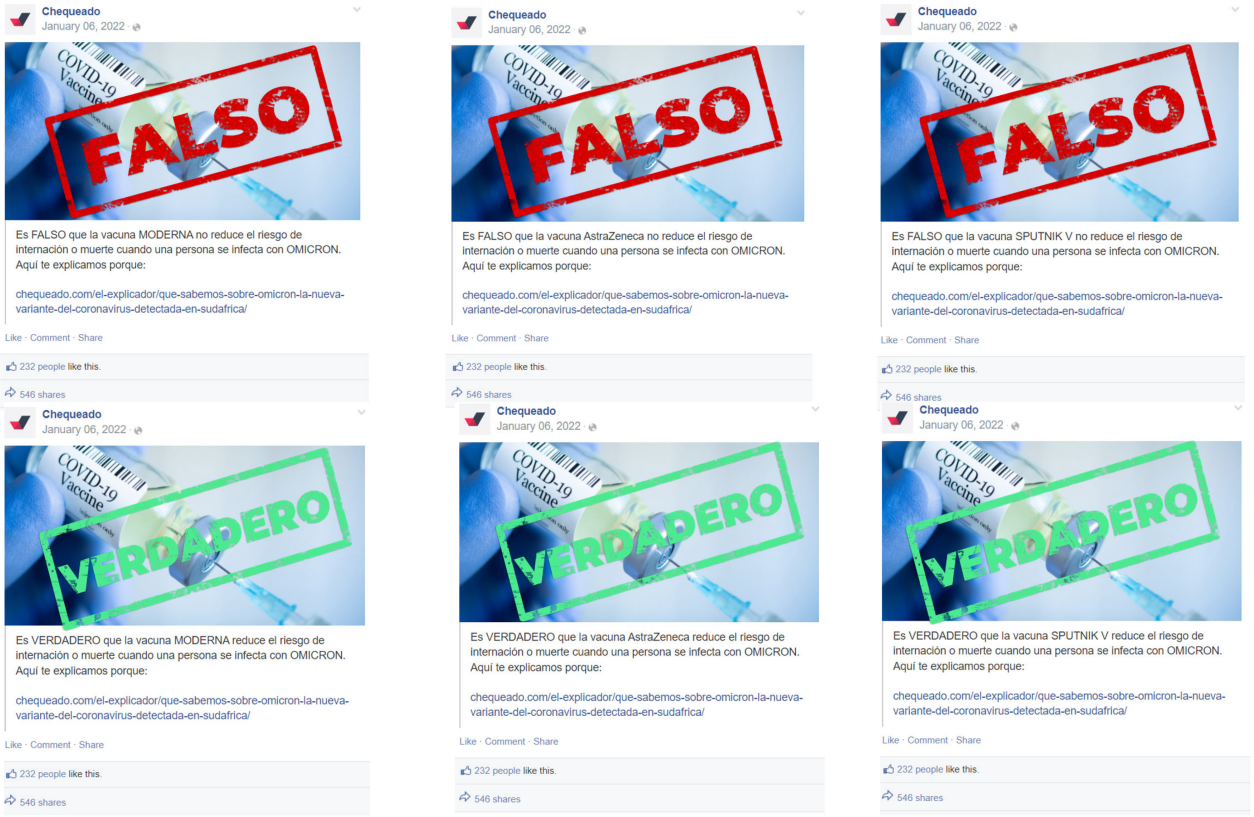


**Figure 1** TRUE and FALSE treatments for each of the vaccines: AstraZeneca, Sputnik V, and Moderna. The TRUE and FALSE statements are semantically identical but differ in their cognitive accessibility and their valence charge. Both the TRUE and FALSE adjudications are factually correct.

After exposure, we ask respondents to indicate if they would like, share, and/or comment on the Facebook post, allowing for multiple choice responses, and with an explicit "ignore" option that is exclusive if selected. Second, we ask to indicate how the post made them feel, with a list that includes Ekman's six basic emotion categories (Ekman and Friesen, 1971): fear, anger, joy, sadness, disgust, and surprise. We also include a second positive category, optimism. We allow multiple responses except for the alternative "indiferent", which is exclusive if selected.

Each of the instruments (the Facebook treatments, the sharing behavior, and the affective response) are presented to survey respondents in the same flow order. We also register the time-to-read (the elapsed time watching the post), the time-to-react (the elapsed time to the behavior question), and the time-to-feel (the elapsed time to the self-reported emotional reaction). We include a variety of demographic, political, and COVID-19 risk controls.

## Results

Table 1 reports the sharing and affective responses to the confirmation and refutation frames and report difference of means tests to assess the overall effect.[2] The same content framed as a confirmation of the accurate statement results in rates of engagement almost twice as large for all three vaccines, increasing from 0.192 (19.2% of likes, shares, and comments together) to 0.372 (37.2%). The largest effect is on the "like" response, which increases three-fold, from 0.084 to 0.243, a difference of 0.158 that is statistically significant at the $p < 0.01$ level. The effect of the confirmation frames on sharing is more modest, increasing from 0.076 to 0.114, close to a 50% increase. Finally, the effect of the confirmation frame on the propensity to comment is not statistically significant.

[2]Full models and robustness checks are reported in the Supplemental Information File to this article.

The effect of the confirmation and refutation frames on the reported emotion is both interesting and consistent with the expectations. Individuals treated to the confirmation frame report a higher frequency of "joyful" and "optimistic", significant at the $p < 0.01$ level. The four-fold increase in reported joy and six-fold increase in optimism is noteworthy. By contrast, the refutation frame was largely associated with more negative emotions, with particularly large estimated differences in "anger", "disgust", and "stress". Interestingly, the rate of respondents that selected the "indiferent" option is almost identical, representing close to 40% of the confirmation and refutation samples.

Figure 2 visually describes the mean engagement rates ("like", "share", "comment") of the confirmation (TRUE label) and refutation (FALSE label) frames. Figure 2 shows no statistically significant difference that would be explained by the brand of the vaccine. The AstraZeneca, Sputnik V, and Moderna rotations yield statistically similar results.

Table 2 presents results only for the "engage" variable, with separate model estimates for the restricted model, Table 2 (1), and unrestricted models with demographic, political, and incidence of the disease variables. The controlled models in Table 2(2)(3)(4) yield almost identical coefficients, which is expected given that treatments were fully randomized. The SIF file to this article shows that the treatment samples were balanced and provides readers with robustness checks to verify each of the results.

**Table 1** Difference of Means between the *confirmation* and *refutation* frames

| Variable | False | True | Diff | P-value |
|---|---|---|---|---|
| *Reactions* | | | | |
| Engage | 0.192 | 0.372 | 0.180*** | 0.000 |
| Like | 0.084 | 0.243 | 0.158*** | 0.000 |
| Share | 0.076 | 0.114 | 0.038*** | 0.001 |
| Comment | 0.041 | 0.048 | 0.006 | 0.450 |
| *Emotions* | | | | |
| Angry | 0.163 | 0.044 | -0.119*** | 0.000 |
| joyful | 0.017 | 0.096 | 0.079*** | 0.000 |
| Disgusted | 0.193 | 0.076 | -0.117*** | 0.000 |
| Optimistic | 0.059 | 0.331 | 0.272*** | 0.000 |
| Stressed | 0.126 | 0.060 | -0.066*** | 0.000 |
| Sad | 0.059 | 0.024 | -0.035*** | 0.000 |
| Fearful | 0.045 | 0.028 | -0.017** | 0.023 |
| Indifferent | 0.441 | 0.414 | -0.027 | 0.186 |

*Note*: Robust standard errors in parentheses .
Full set of models in the SIF file to this article.
*** p < 0.01, ** p < 0.05, p < 0.1

Figure 3 shows interesting partisan heterogeneity in "likes" when treating respondents with the confirmation frame and no statistically significant partisan differences in the rejection frames. A higher preference was expected of voters of the incumbent administration of Alberto Fernandez, who actively supported quarantine measures early in the pandemics and made vaccination a policy priority. While partisan polarization is high in Argentina, most political actors supported widespread vaccination and partisan conflict centered on the quality of the government's response to the pandemics. Both the leading opposition coalition voters (Cambiemos) and independents voters (blanco) shared the confirmation frames TRUE at higher rates than the refutation frames.

**A Rejection of the Cognitive Difficulty Hypothesis**

We find no evidence that of a higher cognitive burden for the FALSE frame. Three findings are relevant: first, we find no association between the level of education and the confirmation and

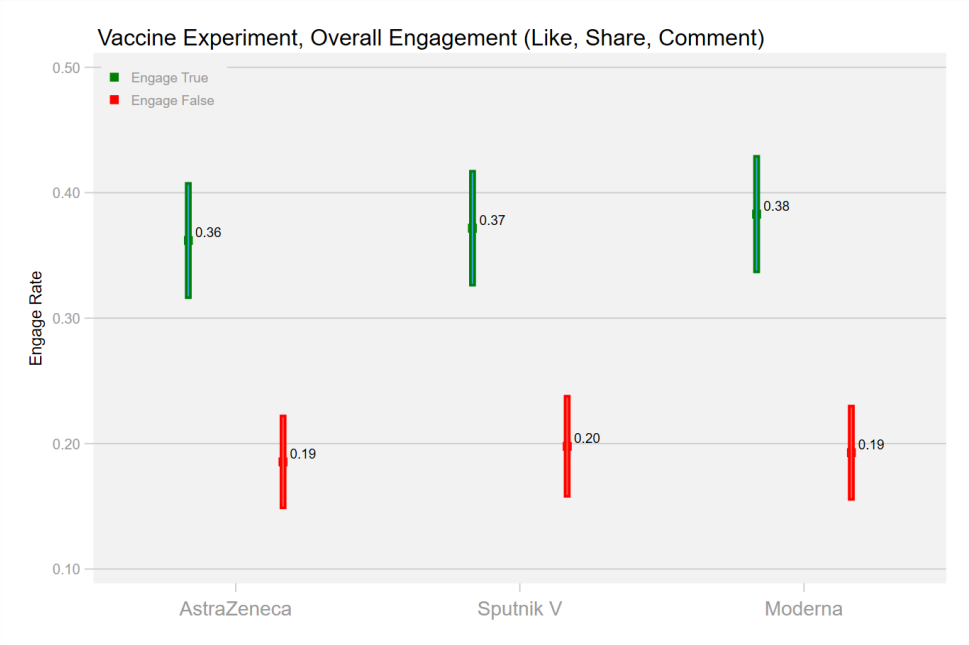


**Figure 2** Overall engagement (like+share+comment) using the confirmation and refutation frames, TRUE or FALSE alternatively. Separate means are presented for each of the vaccines: AstraZeneca, Sputnik V, and Moderna. The TRUE and FALSE statements are semantically identical but differ in their cognitive accessibility and their valence charge. Both the TRUE and FALSE adjudications are factually correct.

refutation frames. Therefore, more educated respondents that should more easily overcome the cognitive burden of refutation frames are indistinguishable from the less educated respondents. Second, there is almost no difference in the time that respondents spent reading the confirmation and refutation frames. Therefore, if there was a higher cognitive burden this did not increase the cognitive effort by the respondents. Finally, and more importantly, an increase in reading time was associated with a statistically significant increase in reported likes ($p < .05$).

This last result, described in Figure 4, is particularly revealing. A faster responses reduces the magnitude and statistical significance between the frames. Indeed, engaging with the TRUE frame increases the longer the time respondents spent reading the Facebook post. Therefore, a more careful reading of the post increases the likelihood that the confirmation frame will be "liked" and "shared".

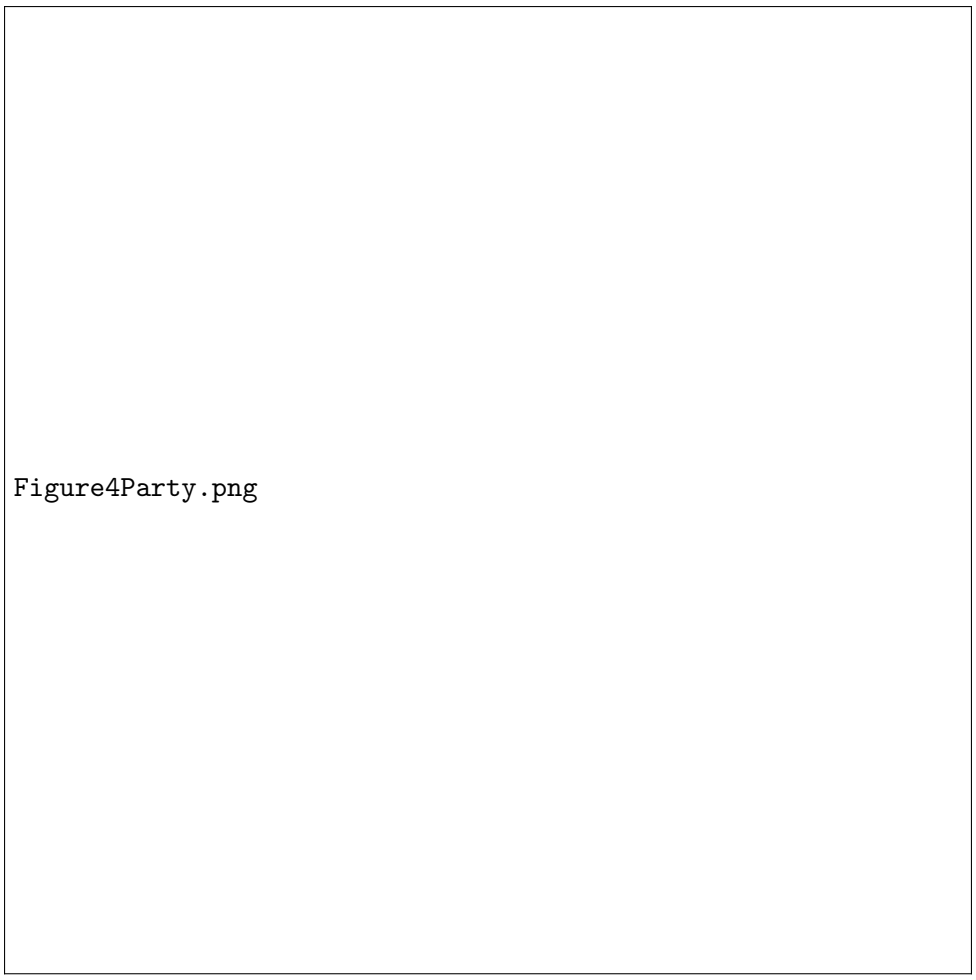


**Figure 3** Like and share rates of the confirmation and refutation frames, TRUE or FALSE. Separate means for incumbent party (Frente de Todos, FdT), the opposition (Cambiemos), and the independent voters (Blanco vote).

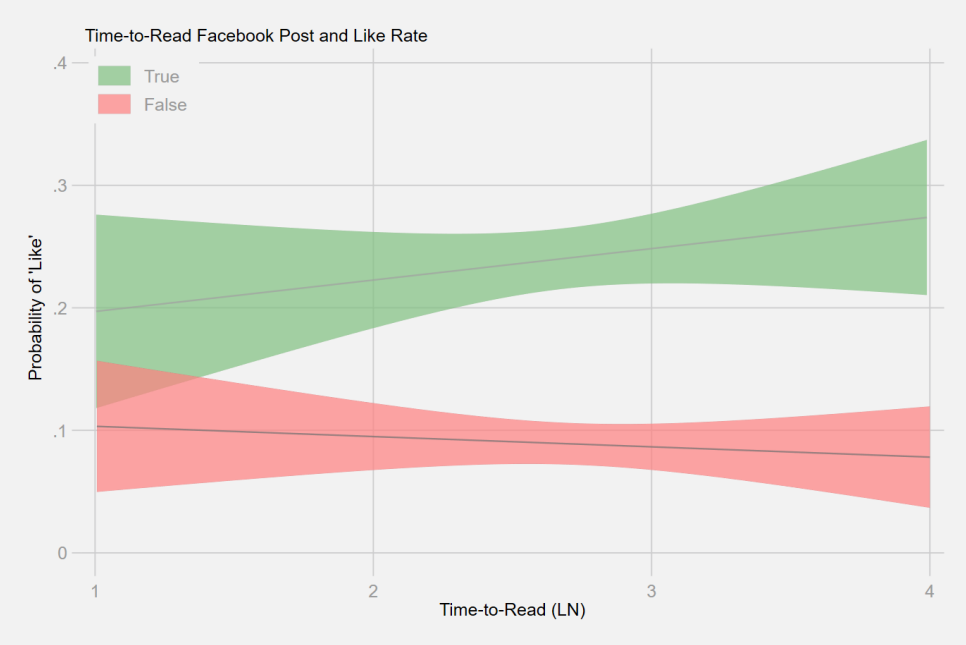


**Figure 4** Like rate and Time-to-Read the Facebook Post. Longer reading times are associated with larger differences in the response to the confirmation and refutation frames.

## Discussion

Results from our survey support a higher intent to "like" and "share" confirmation frames. Consistent with our preregistered hypotheses, the effect of the confirmation frame is positive and statistically significant at $p < 0.01$. Further, the direction of the emotional responses indicates that this is the result of a different interpretation of the confirmation and refutation frames. We speculate that, although the two frames are semantically equivalent, confirmation frames center the attention of the reader on the health benefits of the vaccine while refutation frames focus the attention of the reader on the misinformation event.

A valence driven interpretation of the results is given further support by the rejection of the cognitive burden hypothesis. We find no evidence that rates of liking or sharing result from difficulties in understanding the confirmation and refutation frames. There is no significant difference in the mean processing time for each frame. Further, we document an increase in

likes and shares of the confirmation frame as time increases. Given that the confirmation and refutation time-to-read is similar, but that longer reading times increase the probability of liking and sharing the confirmation frame, the only reasonable explanation is that higher understanding increases the positive valence of the confirmation frame.

The results of our experiment have important policy consequences. Fact-checkers interested in increasing exposure for their posts will be better served using the confirmation frame more frequently. An analyses of the use of TRUE frames compared to FALSE frames among 22 fact checkers in Latin America showed that refutation frames are four times more likely to be used. Further, some fact-checkers only use refutation frames, both reducing the level of exposure of their corrections and in all likelihood increasing the negative valence stock in social media.

Findings in this paper also provide evidence that the effect of confirmation and refutation frames is independent from other demographic, partisan, and health associated moderators of fact-checking sharing. The emphasis on the negative partisan effects of misinformation often times obscure that negative and positive valence charges in health messages are not only the result of our partisan predispositions. Different editorial strategies may be selected to frame a correction as a contribution to the overall endowments of correct information that is present in social media or as a contribution to the overall endowment of polarized content. The standard use of the label "FALSE" may be interpreted not only as a warning on toxic content but also as a reminder to readers that social media is highly polarized. Attention may be deflected from the important health issues and directed instead to the partisan conflict that underlies it.

# Methods

### Survey information

Our survey experiment was deployed in the second wave of the Argentine National Election Survey in February of 2022. The survey was coordinated by the Interdisciplinary Lab for Computational Social Science (iLCSS) at the University of Maryland, College Park and builds on the prior collaboration between the iLCSS and the Argentinian Fact-Ckecking Agency Chequeado on the moderators of factchecking corrections on social media.

The Argentine National Election Survey was deployed online by the pooling firm Netquest [3]. The sample included XXXX adult respondents from the 24 Argentinean provinces, stratified by gender, age, and education to match current census data. The final survey took a median time to completion of XXX minutes, and in addition to the experiment, asked a battery of political questions.

### Design information

The vaccines experiment uses a two-arm design that exposes respondents to a Tweet that randomly confirms or refutes a clinically correct statement describing the vaccines' efficacy against Omicron. Our design randomly assign half of our sample to read a confirmation of accurate information ("it is TRUE that the Moderna vaccine prevents against the new Omicron variant") and other half to read a refutation of misinformation ("it is FALSE that the Moderna vaccine does not prevent against the new Omicron variant"). In addition to the framing rotation, the

[3]Netquest is a reputable survey company with large global panels of respondents. Netquest's panels are opt-in for respondents, using quota sampling to achieve a nationally representative sample on key demographics, such as age, gender, state, and income. An independent assessment of the quality of Netquest panel against a probabilistic sample was conducted recently by Castarena et al. (2021), finding very small deviations from optimal sampling

design also considers three different vaccines (Sputnik V, Moderna, and AstraZeneca). These were the three vaccines with largest coverage in Argentina at the time of the experiment. In conclusion, respondents are randomly assigned to one of six treatment groups (i.e. confirmation or refutation frames for Moderna, AstraZeneca, or Sputnik V).

We summarize here all the sequence of the experiment. We first expose respondents to either a TRUE or FALSE statement about the efficacy against Omicron of the Sputnik V, AstraZeneca, or Moderna vaccines. Then, we measure time-to-read for the correction by the time they take to move to the next online survey page. This measure is used in our results. Next, we then ask respondents a behavioral question. Specifically, we ask if they would "like", "share", "reply", or "ignore" the facebook post. Finally, we ask respondents an emotional question. More specifically, we ask how they felt after reading the adjudication.

Using the random assignment of the experiment, all our statistical models discussed in the paper relies on simple two-tailed mean tests and simple linear models using ordinary least-square estimators. Below, we describe the survey variables used in the analysis.

**Variable definitions**

#### 0.0.1 Dependent variables

- **Engagement.** After seeing the vignette, respondents are asked to choose one or more reactions: Like, retweet (or share), reply, or ignore. Each reaction is treated as a dependent variable by itself. In addition, there is an indicator variable (engage) for the selection of at least one active reaction (like, retweet, or reply) by the respondent.

- **Emotions.** After seeing the vignette, respondents are asked if the publication aroused any of the following emotions in them: Anger, contempt, disgust, optimism, stress, sadness,

fear, or indifference. Respondents can mark more than one option. Each emotion is associated with one indicator variable and is treated as a single dependent variable.

### 0.1 Main independent variables

- **True/False framing.** Binary variable indicating if the statement in the vignette was formulated in a confirmation framework ("It is TRUE that X vaccine REDUCES hospitalization or death risk when a person is infected with OMICRON") or refutation framework ("It is FALSE that X vaccine DOES NOT REDUCE hospitalization...").

- **Time to read.** Time in seconds spent by the respondent at the vignette screen before moving on to the question on engagement.

### 0.2 Political control variables

- **Vaccine mentioned in the vignette.** Set of indicators for the vaccine brand mentioned in the vignette: AstraZeneca, Moderna, or Sputnik V.

- **Partisan attachment.** Set of binary variables indicating vote intention in hypothetical presidential elections: *Frente de Todos* (center-left ruling party), *Juntos por el Cambio* (center-right opposition party), and *voto en blanco* (none of the above).

### 0.3 Demographic control variables

- **Age.** Set of staggered indicator variables for six age groups: 18 to 25 years old, 26 to 35 years old, 36 to 45 years old, 46 to 55 years old, 56 to 65 years old, and more than 65 years old.

- **Sex.** Binary variable indicating if the respondent is a woman.

- **Education.** Set of indicator variables for the highest level of education attained (completed or incomplete): Primary, secondary, university (undergraduate), or graduate level.

- **Employment status.** Binary variable indicating if the respondent is employed at the time of answering the questionnaire.

- **Employment category.** (Only for employed respondents) Set of binary variables indicating the employment category: Private sector employee, federal, province, or municipal public employee, self-employed, business owner with two or more employees, or other.

### 0.4 COVID-19 control variables

- **Vaccination status.** Set of indicator variables for the number of COVID-19 vaccine doses received: None, one, two, or three or more.

- **COVID-19 status.** Binary variable indicating if the respondent ever got COVID-19. There is also an indicator for respondents who do not know if they had got infected or not.

- **COVID-19 sequels.** (Only for those who got infected) Binary variable indicating if the COVID left health sequels (including physical and emotional consequences) to the respondent.

### Ethics

Human Subjects and Ethics approval was granted by the University of Maryland Institutional Review Board on XXXXX. The project approval is registered under the identification code XXXXX XXXXX. Consent was requested at the beginning of the survey and a disclaimer provided respondents with information on how to contact the researchers or IRB if needed.

## References


Arechar, A. A., Allen, J. N. L., Cole, R., Epstein, Z., Garimella, K., Gully, A., Lu, J. G., Ross, R. M., Stagnaro, M., Zhang, J., et al. (2022). Understanding and reducing online misinformation across 16 countries on six continents.

Aruguete, N., Bachmann, I., Calvo, E., Valenzuela, S., and Ventura, T. (2022). Truth be told: Cognitive moderators of selective sharing of fact-checks on social media. *iLCSS*.

Aruguete, N. and Calvo, E. (2018). Time to #Protest: Selective Exposure, Cascading Activation, and Framing in Social Media. *Journal of Communication*, 68(3):480–502.

Bode, L. and Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through related stories functionality in social media. *Journal of Communication*, 65(4):619–638.

Christensen, K. R. (2020). The neurology of negation: fmri, erp, and aphasia. In *The Oxford handbook of negation*, pages 725–739. Oxford University Press Oxford.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., et al. (2020). Real solutions for fake news? measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, 42(4):1073–1095.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., and Lazer, D. (2019). Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kaup, B., Lüdtke, J., and Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, 38(7):1033–1050.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., et al. (2018). The science of fake news. *Science*, 359(6380):1094–1096.

Porter, E. and Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in argentina, nigeria, south africa, and the united kingdom. *Proceedings of the National Academy of Sciences*, 118(37).

Shin, J. and Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, 67(2):233–255.

Van Der Linden, S., Maibach, E., Cook, J., Leiserowitz, A., and Lewandowsky, S. (2017). Inoculating against misinformation. *Science*, 358(6367):1141–1142.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Walter, N., Cohen, J., Holbert, R. L., and Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3):350–375.

## Acknowledgements

We thank Matias Guizzo Altube, Felipe Gonzalez Alzaga, and Carolina Verena Franco for excellent research assistance.

## Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.