

Reducing Misinformation:

The Role of Confirmation Frames in Fact-Checking Interventions

Natalia Aruguete

Flavia Batista

Ernesto Calvo

Matías Guizzo Altube

Carlos Scartascini

Tiago Ventura

Reducing Misinformation:

The Role of Confirmation Frames in Fact-Checking Interventions

Natalia Aruguete*

Flavia Batista**

Ernesto Calvo**

Matías Guizzo Altube***

Carlos Scartascini***

Tiago Ventura****

* Universidad Nacional de Quilmes

** University of Maryland

*** Inter-American Development Bank

**** New York University

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Reducing misinformation: the role of confirmation frames in fact-checking
interventions / Natalia Aruguete, Flavia Batista, Ernesto Calvo, Matías Guizzo Altube,
Carlos Scartascini, Tiago Ventura.

p. cm. — (IDB Working Paper Series ; 1363)

Includes bibliographic references.

1. Misinformation-Latin America. 2. Misinformation-Caribbean Area. 3. Fake news-
Latin America. 4. Fake news-Caribbean Area. 5. Social media-Social aspects-Latin
America. 6. Social media-Social aspects-Caribbean Area. 7. Coronavirus infections-
Social aspects-Latin America. 8. Coronavirus infections-Social aspects-Caribbean
Area. I. Aruguete, Natalia. II. Batista, Flavia. III. Calvo, Ernesto. IV. Guizzo, Matías.
V. Scartascini, Carlos G., 1971- VI. Ventura, Tiago. VII. Inter-American Development
Bank. Department of Research and Chief Economist. VIII. Series.
IDB-WP-1363

<http://www.iadb.org>

Copyright © 2023 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose, as provided below. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Following a peer review process, and with previous written consent by the Inter-American Development Bank (IDB), a revised version of this work may also be reproduced in any academic journal, including those indexed by the American Economic Association's EconLit, provided that the IDB is credited and that the author(s) receive no income from the publication. Therefore, the restriction to receive income from such publication shall only extend to the publication's author(s). With regard to such restriction, in case of any inconsistency between the Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives license and these statements, the latter shall prevail.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



Abstract

Previous research has extensively investigated why users spread misinformation online, while less attention has been given to the motivations behind sharing fact-checks. This paper reports a four-country survey experiment assessing the influence of *confirmation* and *refutation* frames on engagement with online fact-checks. Respondents randomly received semantically identical content, either affirming accurate information (“It is TRUE that p ”) or refuting misinformation (“It is FALSE that *not* p ”). Despite semantic equivalence, confirmation frames elicit higher engagement rates than refutation frames. Additionally, confirmation frames reduce self-reported negative emotions related to polarization. These findings are crucial for designing policy interventions aiming to amplify fact-check exposure and reduce affective polarization, particularly in critical areas such as health-related misinformation and harmful speech.

JEL classifications: D83, D91

Keywords: Misinformation, Fact-checking, Social media

We thank Laura Zommer, Olivia Sohr, Ariel Riera, and Leticia Small from *Chequeado* for their suggestions to the design and the materials used for this research project. We also thank Felipe Gonzalez Alzaga, Carolina Batista, and Carolina Verena Franco for excellent research assistance. Previous versions of this article benefited from comments from colleagues at Emory University, the University of Chicago, the University of Maryland, Facebook, and Twitter. Human Subjects and Ethics approval was granted by the University of Maryland Institutional Review Board prior to the implementation of surveys in each country. The project approvals are registered under the identification code IRB 1825785 (1825785-1 to 1825785-8).

1 Introduction

Fact-checking is the first line of defense against misinformation (Bode and Vraga, 2015; Del Vicario et al., 2016; Lazer et al., 2018; Van Der Linden et al., 2017). 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 influence people’s discernment of misinformation and *nudge* users to update their beliefs after correction, whether in survey experiments or field experiments, and across different cultural contexts (Arechar et al., 2022; Bode and Vraga, 2015; Clayton et al., 2020; Porter and Wood, 2021). The effect of fact-checking interventions extends over time, with minimal evidence of backfire effects from exposure to fact-checking corrections (Nyhan et al., 2020; Swire-Thompson et al., 2020).

To curb the spread of misinformation, fact-checkers can employ two distinct framing strategies: they can either publish **confirmation frames** that replace misinformation with accurate information, or they can publish **refutation frames** that warn social media users about content tagged as misinformation. Choosing *confirmations* provides users with factually accurate content they can share with peers. Opting for *refutations* allows fact-checkers to decrease the sharing of inaccurate, misleading, or false content. The effectiveness of increasing “good” content versus reducing “bad” content has not been experimentally tested. In this paper, we evaluate the impact of *confirmation* and *refutation* frames on the sharing behavior of social media users.

The lack of studies measuring the impact of *confirmation* (*TRUE*) and *refutation* (*FALSE*) frames is surprising, given the central role content labeling plays in fact-checking interventions. 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). The decision to intervene using *confirmation* or *refutation* frames is an editorial choice that is independent of the source material (Vosoughi et al., 2018).

This paper presents experiments conducted in four different countries to assess the effect of *confirmation* and *refutation* frames on the sharing behavior of social media users. Our experiments expose nationally representative samples of Argentine, Brazilian, Chilean, and Colombian respondents to edited Facebook posts framed as a confirmation of accurate information or a refutation of misinformation. The experiment rotates the confirmation and refutation frames, the choice of labels (labeled vs. unlabeled), and the type of vaccine (Moderna, AstraZeneca, and Sputnik V).¹ The empirical analysis and robustness checks include several control variables (i.e., socio-demographic, attitudinal, and health status variables) and validation checks (i.e., processing time and pseudo-placebo treatment).

Our primary outcome measures the decision to engage (i.e., “like,” “share,” and “comment”) with the fact check and the self-reported affective response to the fact-checking post. Our hypotheses, pre-registered at <https://osf.io/> prior to the collection of the data, posit that respondents will engage more with confirmation frames than refutation frames [hypothesis 1 (H1)]. We propose this effect to be independent of other factors prompting engagement with a correction, such as cognitive difficulty, cognitive congruence, and partisan attachment. Our primary hypothesis stems from two theoretical mechanisms: the heavier cognitive burden of refutation frames and the positive valence charge associated with confirmation frames. We offer specific hypotheses and dedicated tests for each mechanism.

First, negation is known to impose a heavier cognitive load (Christensen, 2020). Research in cognitive linguistics and cognitive psychology has documented differences in processing semantically equivalent positive or negative statements. Kaup et al. (2006) show that individuals

¹The variation in vaccines only takes place in Argentina.

are faster to process statements 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” than for “the umbrella was not open.” Indeed, this 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 networks, a higher cognitive burden could conceivably deter a swift, automatic, and affective response (Arugete and Calvo, 2018; Kahneman, 2011), leading to more evaluative sharing behavior. We hypothesize that refutation frames exert a higher cognitive burden on respondents, thus resulting in longer reading times [hypothesis 2 (H2)] that curtail sharing.

Second, we expect that the confirmation of pro-attitudinal beliefs will carry a positive valence charge compared to the refutation of a counter-attitudinal belief. A standard sentiment analysis using state-of-the-art RoBERTa (Loureiro et al., 2022) shows that “It is true that vaccines are effective” is classified as *Positive* (i.e., Cardiff scores are *Positive*: 0.782, *Neutral*: 0.209, *Negative*: 0.009). Meanwhile, “It is false that vaccines are not effective” is classified as *negative* (i.e., Cardiff scores are *Positive*: 0.024, *Neutral*: 0.278, *Negative*: 0.698). This is because the words “true” and “false” function not only as Boolean operators but also convey positive and negative connotations in social conversation.

In addition, there are also social differences in being recognized as someone who tells the truth or someone who lies, which also carry a positive valence charge. Confirmation statements such as “it is TRUE” convey that the content is socially acceptable and less likely to expose users to public scrutiny and criticism.² Confirmation frames communicate greater social acceptability and widespread consensus with published content. Refutation frames, in contrast, suggest that there are dissenting opinions and raise the potential for conflict. That is, refutation frames

²Tetlock (2002) coins the term “intuitive politician” to describe the behavior of risk-averse subjects who seek to preserve their reputation by aligning themselves with socially accepted positions. “People behave like intuitive politicians when they seek to maintain a positive reputation or fulfill the social duties for which they are accountable” (Margolin et al., 2018).

suggest that there are at least some individuals or groups with competing beliefs (Tetlock, 2002). Therefore, statements framed as confirmations will have a positive valence charge that is independent of the pro- or counter-attitudinal preferences for the denoted content in the message. We hypothesize that confirmation frames will elicit positive emotional reactions and refutations will elicit negative ones [hypothesis 3 (H3)].

To sum up our pre-registered hypotheses, we anticipate the statement “it is TRUE that p” to enhance engagement compared to “it is FALSE that not p” [hypothesis 1 (H1)], because the former is both cognitively simpler to process [hypothesis 2 (H2)], and because *TRUE* carries an inherent positive valence charge [hypothesis 3 (H3)]. Conversely, refutation statements are cognitively challenging, and sharing refutation messages aligns one with an in-group social media user at odds with an out-group user’s beliefs.

2 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. Crucially, the experiment did not spread misinformation to participants; both the *confirmation* and the *refutation* frames communicated that vaccines are effective against the Omicron variant. In the Argentine version of the experiment, three different vaccines (Sputnik V, Moderna, and AstraZeneca) were taken into account to test for differences in perceived vaccine quality. For each vaccine, participants were exposed to *confirmation* and the *refutation* frames. In Chile, Brazil, and Colombia, two distinct designs were employed, presenting *confirmation* and *refutation* treatments either with explicit labels or without labels (see Figure 1 for the treatments implemented in Colombia—the complete set of treatments is provided in the Supplementary Information File (SIF)).³

Additionally, we introduced confirmation and refutation frames unrelated to our health cor-

³See Figures S1, S2, S3, and S4.

rection and devoid of any correlation with political preferences. This pseudo-placebo treatment measures the independent valence charge associated with the use of the words “true” and “false” in a post about dogs.

In all four countries, simple randomization was implemented, with respondents having equal chances of being assigned to each treatment (*confirmation* or *refutation* of the vaccines or dog treatments) and to each of the design alternatives (label, no-label, and, in the case of Argentina, vaccine type).⁴

After exposure to the treatments, respondents were asked to indicate whether they would “like,” “share,” and/or “comment” on the Facebook post. The response format allowed for multiple selections, with an explicit “ignore” option that was exclusive if chosen. Additionally, participants were asked to self-report their emotional response to the post, choosing from a list that included Ekman’s six basic emotion categories: *fear*, *anger*, *joy*, *sadness*, *disgust*, and *surprise*, as well as an additional positive category, *optimism*. Multiple responses were allowed, except for the alternative *indifferent*, which was exclusive if selected.

The sequence of presentation (the Facebook treatments, the sharing behavior, and the emotional response) remained the same for all survey respondents. Additionally, the researchers recorded the time-to-read (the elapsed time spent viewing the post), the time-to-react (the elapsed time before responding to the behavior question), and the time-to-feel (the elapsed time before reporting an emotional reaction). The survey collected additional information to allow the inclusion of various demographic, political, and COVID-19 risk factors in the empirical analysis.

⁴Table S1 through Table S4 present summary statistics and balance across the treatments.



Figure 1 Images of the *Confirmation* (“It is TRUE that p”) and *Refutation* (“It is False that not p”) treatments used in Colombia. Left images without labels. Right images with labels. The confirmation and refutation frames are semantically equivalent but differ in their cognitive accessibility and their valence charge. All four treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*. The designs for each country, and the placebo, are reported in the Supplemental Information File to this article.

3 Empirical Methods

Survey Information

The survey experiments were conducted in Argentina in February of 2022, Chile in November of 2022, Brazil in December of 2022, and Colombia in March of 2023. The surveys were designed by the Interdisciplinary Lab for Computational Social Science (iLCSS) at the University of Maryland, College Park, in collaboration with the Fact-Checking Agency Chequeado.

All four surveys were administered online by the polling firm Netquest.⁵ The sample comprised 12,000 adult respondents from Argentina, Brazil, Chile, and Colombia, with stratification based on gender, age, and education in accordance with current census data. The survey took a median time of 22 minutes to complete. In addition to the experiment, it included a battery of socio-demographic, attitudinal, and political questions.

Design Information

The “vaccine” experiments use a two-arm design that exposes respondents to one of two equivalent statements that *confirm* the efficacy of the vaccine or refute their inefficacy. The design randomly prompts respondents to read either the confirmation statement “*It is TRUE that the new #VacunaBivalente is effective against the Omicron variant*” or the refutation of the corresponding misinformation “*It is FALSE that the new #VacunaBivalente is not effective against the Omicron variant.*” In Argentina, the brand of the vaccine (Sputnik V, Moderna, and AstraZeneca) is rotated. In Brazil, Chile, and Colombia, the use of labels is rotated, and a pseudo-placebo treatment about dogs is included.⁶

⁵Netquest is a reputable survey company with large global panels of respondents. Netquest panels opt-in respondents, using quota sampling to achieve a nationally representative sample on key demographics, such as age, gender, population, and income. An Independent assessment of the quality of Netquest panels compared with a probabilistic sample was recently published by (Castorena et al., 2023), finding very small deviations from optimal sampling

⁶The treated individuals are well-balanced as shown in Table S1 to S4 in the SIF.

The flow of the experiment is as follows. First, respondents are exposed to either a *confirmation* or *refutation* frame. The time respondents spend reading the statement (time-to-read) is measured, beginning with the image loading and ending when the respondent progresses to the next page of the online survey. The second page asks respondents if they would “like,” “share,” “comment,” or “ignore” the Facebook post. The time-to-respond is again measured. Finally, on the third page, respondents are asked to self-report their emotional reaction to the question.

The statistical models utilized in the paper employ simple two-tailed mean tests. The conditional effects of time variables and other socio-economic indicators are further assessed using general linear regression models as well as ordinary least-square models.

Variable Definitions

Dependent Variables

- **Engagement.** After seeing the Facebook Post, respondents are asked to “like,” “share,” “comment”, or ignore it. Each reaction is treated as a dependent variable. 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 Facebook Post, respondents are asked if the publication elicited any of the following emotions: 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.

Treatment Variables

- **True/False framing.** Binary variable indicating if the statement is a *confirmation* or *refutation* frame (“It is TRUE that p”) or refutation framework (“It is FALSE that not p”).

- **Argentina: Brand of the vaccine.** Set of indicators for the vaccine brand mentioned in the vignette: AstraZeneca, Moderna, or Sputnik V.
- **Brazil, Chile, and Colombia: Explicit label used in the treatment.** A categorical variable that indicates if the label was included or not.

Control Variables

- **Time to read.** Time in milliseconds (log) spent by the respondent.
- **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).
- **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.
- **Vaccination status.** Set of indicator variables for the number of COVID-19 vaccine doses received: None, one, or two or more.
- **COVID-19 status.** Binary variable indicating if the respondent ever got COVID-19.

4 Results

Confirmation frames led to systematically higher engagement than *refutation* frames across the four countries involved in the study. Detailed results are shown in Figure 2, and full regression tables are available in the Supplemental Information File.⁷ Focusing on the *engagement* category (as represented by the first set of bars for each country in Figure 2), it was evident that the confirmation frame significantly increased overall engagement with the correction, affirming hypothesis 1 (H1). For example, in Argentina, engagement rose from 0.189 (or 18.9% of combined likes, shares, and comments) to 0.371 (or 37.1%), which is a significant positive difference of 18.2 percentage points (effectively a two-fold increase). Differences in engagement for Brazil, Chile and Colombia amounted to 13, 15, and 14 percentage points, respectively.

When examining the individual components of engagement, similar trends were observed. In all four countries, the “like” category showed the highest differences. For Argentina and Chile, confirmation frames produced a three-fold higher proportion of likes compared to refutation frames. For Brazil and Colombia, the difference was about twice as large.

The impact of the confirmation and refutation frames on reported emotion [hypothesis 3 (H3)] is consistent with our expectations. As illustrated in Figure 3, individuals who were exposed to the confirmation frame reported significantly more “joyful” and “optimistic” responses, significant at the $p < 0.01$ level. These differences are quite pronounced, ranging from a more than two-fold increase in reported optimism and joy in Brazil, to more than a five-fold increase in Argentina. In contrast, the refutation frame was primarily linked with negative emotions, such as “anger,” “disgust,” and “stress.” For instance, in Argentina, the refutation frame was four times more likely to induce anger than the confirmation frames. Similarly, the refutation frame was at least twice as likely to elicit anger in the other countries included in the study.

⁷Full models and robustness checks for all four countries are reported in Tables S5, S6, S7, and S8 of the Supplemental Information File accompanying this article.



Figure 2 Regression results on “engagement,” “like,” “share” and “comment” for the four countries. The first red bar for each dependent variable corresponds to the refutation frame. The third green bar corresponds to the confirmation frame. The middle bar shows the difference between the refutation and confirmation frames. When the difference is positive (confirmation frames generate more engagement than refutation) the bar is light green. Light red when the difference is negative. Full regression results are provided in the Supplemental Information File to this article. The summary of the differences in engagement between frames for the four countries including controls is shown in Table 1.



Figure 3 Regression results for the reported emotions for the four countries. The first red bar for each dependent variable corresponds to the refutation frame. The third green bar corresponds to the confirmation frame. The middle bar shows the difference between the refutation and confirmation frames. When the difference is positive (confirmation frames generate more engagement than refutation), the bar is light green. It is light red when the difference is negative. Full regression results are provided in the Supplemental Information File to this article. The summary of the differences in engagement between frames for the four countries including controls is shown in Table 1.

Table 1 presents the results of the differences—depicted as the middle bars in the figures—for all four countries. It emphasizes the variances between the *refutation* and *confirmation* frames, taking into account various control measures.⁸ The results presented in Table 1 underline the consistency of the proposed framing effects, suggesting a notable influence of the confirmation and refutation frames on the outcomes observed.

Across all four countries, Table 1 consistently indicates that the emotional responses to the treatments are more positive for the confirmation frames and more negative for the refutation frames. The confirmation frame elicits feelings of optimism and joyfulness, according to the self-reported emotions of respondents. On the other hand, the refutation frame consistently provokes more negative emotions, such as anger, sadness, stress, fear, and disgust.

Interestingly, even though the refutation frame communicates the same core information as the confirmation frame, it consistently evokes stronger negative emotional reactions. This finding suggests that the manner in which information is presented or framed plays a crucial role in determining its emotional impact on the recipient. The refutation frame seems to incite affective polarization, a phenomenon that is well-documented in the existing literature.

Figure 4 illustrates the findings from the Argentine survey, showcasing the average rates of engagement (i.e., the sum of “like,” “share,” and “comment” rates) for the *confirmation* frame (TRUE label) and the *refutation* frame (FALSE label). These results validate the expectation that the effect of framing is consistent regardless of the specific vaccine brands. The results from testing with the AstraZeneca, Sputnik V, and Moderna vaccines were statistically indistinguishable. This is especially noteworthy given that different vaccines were associated with political decisions, and thereby became ideologically charged, and were believed to have varying degrees of effectiveness. Such consistency emphasizes the robustness of the framing effect. The

⁸The complete set of results can be found in Table S5 through Table S8 in the Supplemental Information File (SIF). The SIF also includes extensive robustness checks, alternative estimates with and without controls, estimates with heterogeneous effects for socio-demographic and attitudinal questions, and heterogeneous effects by party.

Table 1 Difference of Means between the *Confirmation* and *Refutation* Frames

Variable	Argentina	Brazil	Chile	Colombia
<i>Reactions</i>				
Engage	0.188*** (0.018)	0.131*** (0.024)	0.152*** (0.023)	0.147*** (0.024)
Like	0.163*** (0.015)	0.127*** (0.022)	0.171*** (0.018)	0.120*** (0.020)
Share	0.042*** (0.012)	0.001 (0.018)	0.010 (0.017)	0.035* (0.019)
Comment	0.008 (0.009)	-0.002 (0.015)	0.006 (0.010)	0.026** (0.012)
<i>Emotions</i>				
Optimistic	0.278*** (0.015)	0.258*** (0.022)	0.246*** (0.020)	0.226*** (0.022)
Joyful	0.087*** (0.009)	0.106*** (0.018)	0.105*** (0.014)	0.071*** (0.013)
Angry	-0.121*** (0.012)	-0.083*** (0.014)	-0.081*** (0.015)	-0.049*** (0.013)
Sad	-0.035*** (0.008)	-0.102*** (0.014)	-0.023*** (0.009)	-0.049*** (0.011)
Stressed	-0.072*** (0.012)	-0.051*** (0.012)	-0.025* (0.015)	-0.028* (0.015)
Fearful	-0.015* (0.008)	-0.051*** (0.014)	-0.042*** (0.015)	-0.037** (0.019)
Disgusted	-0.121*** (0.014)	-0.112*** (0.014)	-0.046*** (0.015)	0.001 (0.011)
Indifferent	-0.028 (0.020)	-0.029 (0.023)	-0.102*** (0.025)	-0.133*** (0.025)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the survey of the country indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction or emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Full set of models in the SIF file to this article.

Supplemental Information File provides a comprehensive description of the findings for each behavior separately (e.g., “like,” “share,” and “comment”).

Figure 5 visualizes the engagement results from Brazil, Chile, and Colombia, contrasting the effects of treatments with explicit labels to those without labels.⁹ In Brazil and Colombia, the presence of labels produces larger differences in engagement, ranging from 1.5 to 6 percentage points. In Chile, however, higher differences are noted when there are no labels. Intriguingly, across all three countries, individuals are more inclined to engage with negative frames when

⁹Labels refer to the large banners placed over the picture, as depicted in Figure 1.

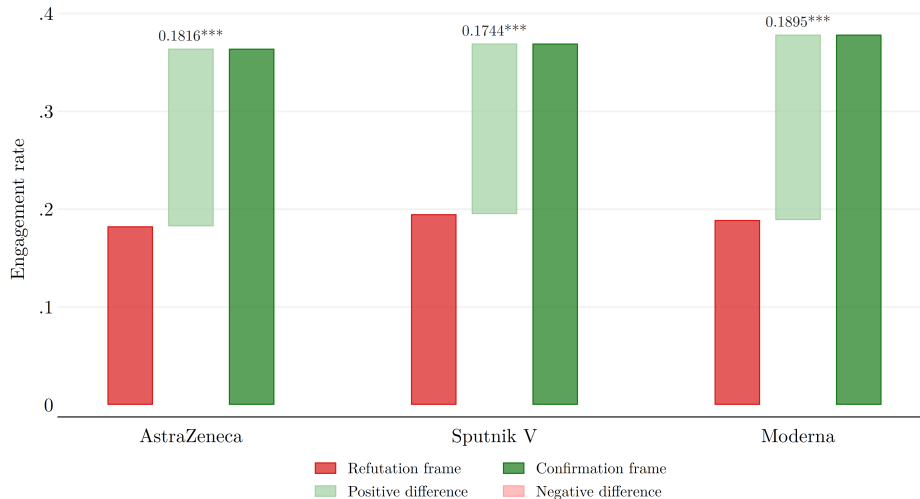


Figure 4 Argentine experiment: Overall engagement (like+share+comment) using the confirmation and refutation frames, TRUE or FALSE alternatively. Separate means are presented for each vaccine brand: AstraZeneca, Sputnik V, and Moderna. The TRUE and FALSE statements are semantically identical but differ in their cognitive accessibility and their valence charge. Both the TRUE and FALSE adjudications are factually correct.

they lack a label.¹⁰

Beyond Vaccines: Dogs Do Not Understand What We Say to Them

In order to examine if the observed effects were particularly influenced by the context of vaccines, a topic that has been heavily politicized in numerous countries, we conducted a distinct exercise to test the robustness of our findings. This experiment involved presenting survey participants with a minimally modified CNN social media post framed either as a confirmation (“True: Dogs do not really understand what we say to them”) or as a refutation (“False: Dogs do not really understand what we say to them”). As can be seen in Figure 6, which uses the treatments in Brazil as an example, the only difference between the two treatments lies in the inclusion of the words “True:” or “False:”. This ensures that the same content is conveyed while maintaining consistency in the semantic meaning and cognitive complexity of the message.

This treatment offers an opportunity to investigate the direct and unconditioned impact of

¹⁰Separate estimates for the components of engagement can be found in the supplemental information file.

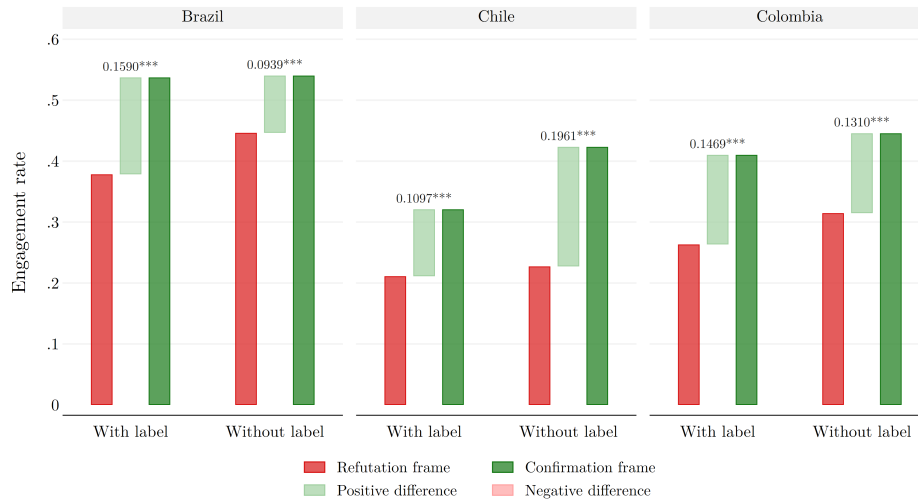


Figure 5 “Engagement” rate using the confirmation and refutation frames, TRUE or FALSE alternatively. Separate means are presented for the treatments with and without explicit labels. The TRUE and FALSE statements are semantically identical but differ in their cognitive accessibility and their valence charge. Both the TRUE and FALSE adjudications are factually correct.

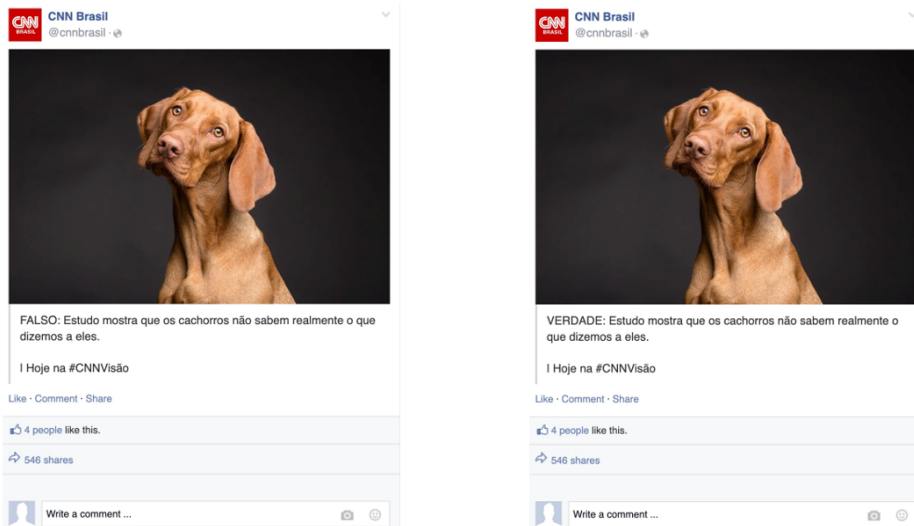


Figure 6 Images of the *confirmation* (“It is TRUE that p”) and *refutation* (“It is False that not p”) pseudo-placebo treatments used in Brazil. The confirmation and refutation frames are semantically equivalent and intended to be equivalent in their cognitive accessibility and their valence charge, changing only the word “True” for “False.” Both treatments conform to the design used by our partner organization in Argentina, *Chequeado*. The pseudo-placebo designs for each country are reported in the Supplemental Information File of this article.

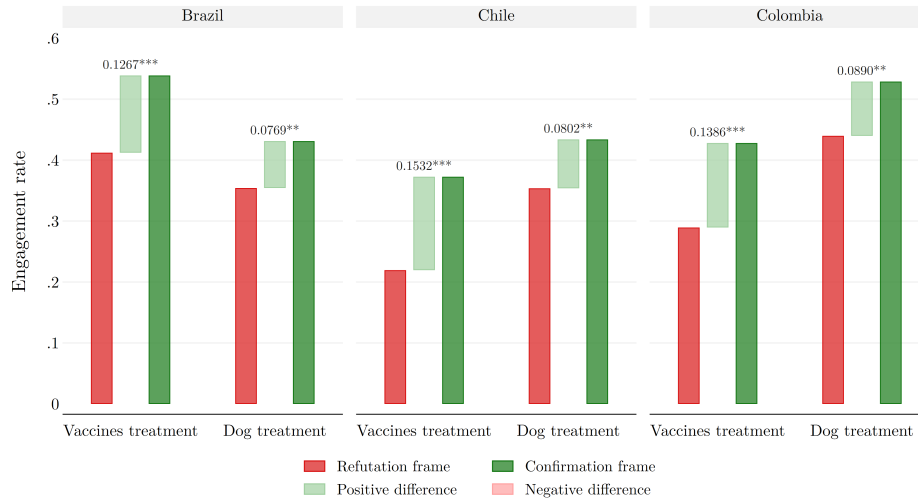


Figure 7 “Engagement” rate using the confirmation and refutation frames, TRUE or FALSE alternatively. Separate means are presented for the dog treatment (pseudo-placebo) and the vaccine treatments (pooling labeled and unlabeled treatments). The TRUE and FALSE statements are semantically identical but differ in their cognitive accessibility and their valence charge. Both the TRUE and FALSE adjudications are factually correct.

the words “True” and “False.” Figure 7 includes the estimates of the “dog” treatment for Brazil, Chile and Colombia.¹¹ The results point to positive and statistically significant effects of the confirmation frame on “engagement.” While these differences are smaller in percentage terms compared to the previous analysis, they still present differences that range between 8 and 9 percentage points, equating to roughly a 20 percent increase in engagement across all three countries. Furthermore, the effects on the “Like” behavior in Brazil and Chile are also positive and statistically significant. This simple exercise shows the power of the “TRUE” and “FALSE” labels on any type of post, indicating the substantial influence on engagement levels.

A Rejection of the Cognitive Difficulty Hypothesis, H2

Our findings provide no evidence to suggest a higher cognitive burden associated with the FALSE frame, as hypothesized in Hypothesis 2 (H2). Two major observations support this conclusion. First, there is no consistent relationship between education level and engagement

¹¹The full results can be found in Table S28 through Table S30 of the Supplemental Information File.

with the confirmation and refutation frames, indicated by the absence of significant patterns across countries. This can be seen in Table S15 of the Supplemental Information File (SIF). The influence of the confirmation frame across varying education levels, both on the propensity to react (Table S24) and the self-reported emotion (Table S24), does not show that more educated individuals are less susceptible to the frames.

Second, we notice no significant decrease in the impact of the confirmation versus refutation frames attributable to the time respondents spent reading the treatments. Contrary to our expectations, longer reading times did not lessen the behavioral and emotional differences between the confirmation and refutation frames. In fact, in Argentina and Brazil, an increase in reading time correlated with a statistically significant rise in reported “likes” for the confirmation frame ($p < 0.05$). The influence of reading time on the overall engagement behavior is illustrated in Figure 8 for respondents from all countries in the vaccines treatments.

This impact of extended exposure time is significant: prolonged exposure to the TRUE frame amplifies the differences in “likes” between the confirmation and refutation frames. Thus, a more thorough reading of the post increases the probability that the confirmation frame will attract a higher “like” rate than the refutation frame. Similar results are reported in Tables S26 and S27 of the Appendix for all countries, with Brazil demonstrating results analogous to those of Argentina, while Chile and Colombia show a more modest positive correlation. In none of the four cases is there a statistically significant decline in the behavioral response gap between the confirmation and refutation frames. Consequently, the increased propensity to share the confirmation frame can be exclusively attributed to its positive valence charge, as per Hypothesis 3 (H3), rather than the cognitive difficulty associated with the refutation frame, as per Hypothesis 2 (H2).

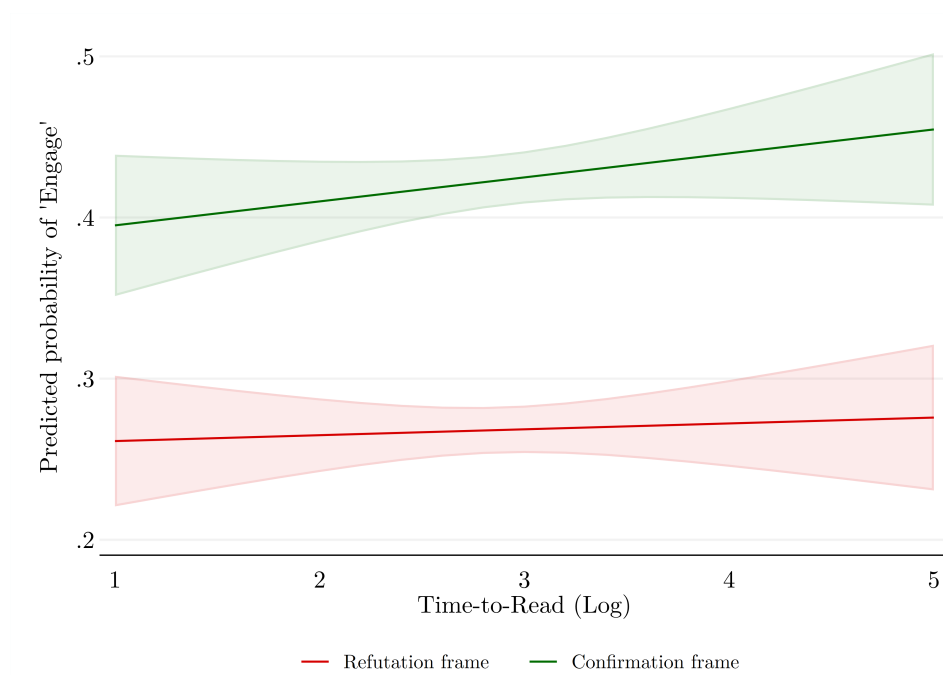


Figure 8 “Engagement” rate and Time-to-Read the Facebook Post in the vaccines treatments. Longer reading times are associated with larger differences in the response to the confirmation and refutation frames. The results refute the cognitive difficulty hypothesis, as higher attention does not reduce the differences between the *confirmation* and *refutation* frames. Probability estimates are obtained from a linear probability model, controlling for socio-demographic characteristics. Shaded area corresponds to the 95% confidence interval.

Other Results: Partisanship, Vaccination Status, and Other Sources of Heterogeneity

In addition to heterogeneity in education and reading time, the Supplementary Information File presents additional exercises where we look at differences according to political affiliation, vaccination status, and other socio-demographic indicators. The differences in engagement remain for individuals in each of these different group categories. Confirmation frames about the vaccines tend to elicit more relative engagement and positive emotions among government supporters in Argentina, where the incumbent government aggressively pursued quarantine and mask mandates. On the other hand, the opposite is true for Brazil and Chile, where anti-COVID policies were more divisive and weakly enforced.¹²

Across the four countries, the differences in engagement and positive emotions tend to be higher for those who were vaccinated (twice or more) than those who were not.¹³ These results suggest that while the effects appear to be fairly universal, variations do exist among different groups, in line with expectations. Therefore, the impact of different fact-checking strategies will not be uniform across all individuals. This indicates that tailoring the framing and the message to suit specific demographics could still be desirable for maximizing the efficacy of the message.

5 Discussion

The results from the four survey experiments lend support to a higher intention to “engage” and “like” fact-checks framed as confirmations compared to the semantically equivalent refutation [hypothesis 1 (H1)]. All four surveys provide evidence that the findings remain robust across a variety of experimental designs, including different brands of the COVID-19 vaccine, with or without the use of labels, and across a diverse range of socio-demographic categories.

Moreover, the observed emotional responses and the absence of an effect related to cognitive

¹²See Tables S16, S17, S18, and S19 in the SIF.

¹³See Table S20 and S21 in the SIF.

effort suggest that this discrepancy arises from distinct interpretations of the confirmation and refutation frames [hypothesis 3 (H3)]. We speculate that, despite their semantic equivalence, confirmation frames draw the reader’s attention towards the health benefits of the vaccine, while refutation frames draw attention to the misinformation event itself.

The rejection of the cognitive burden hypothesis [hypothesis 2 (H2)] further bolsters a valence-driven interpretation of the results. We find no evidence suggesting that rates of liking or sharing stem from difficulties in comprehending the confirmation and refutation frames. Nor is there a significant difference in the mean processing time for each frame. Intriguingly, we observe an increase in “likes” and “shares” for the confirmation frame with prolonged reading times. Given that the reading time is similar for both the confirmation and refutation frames, yet longer reading times increase the probability of liking and sharing the confirmation frame, the only plausible explanation is that a deeper understanding enhances the positive valence of the confirmation frame.

The results of our experiments have significant policy implications. Fact-checkers aiming to expand their posts’ reach would likely benefit from more frequent use of the confirmation frame. Our analysis of TRUE versus FALSE frames usage among 22 fact-checkers in Latin America revealed that refutation frames are four times more likely to be used. Some fact-checkers exclusively use refutation frames, thereby potentially reducing their corrections’ exposure and likely increasing the stock of negative valence content on social media.

The findings in this paper also indicate that the effect of confirmation and refutation frames operates independently of other demographic, partisan, and health-associated moderators of fact-check sharing. The often emphasized negative partisan effects of misinformation can overshadow the fact that negative and positive valence charges in health messages are not solely a result of our partisan predispositions. Fact-checkers can choose different editorial strategies to frame a correction either as a contribution to the overall amount of correct information present

on social networks or as a contribution to the overall stock of polarized content. The standard use of the label “FALSE” can be seen not only as a warning about toxic content but also as a reminder to readers that social media is highly polarized. This may divert attention away from crucial health issues and towards the partisan conflict underlying them.

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. 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.
- Castorena, O., Lupu, N., Schade, M., and Zechmeister, E. (2023). Online surveys in latin america. *PS: Political Science Politics*, 56(2):273–280.
- 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.
- 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.
- Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., and Camacho-Collados, J. (2022). Timelms: Diachronic language models from twitter. *arXiv preprint arXiv:2202.03829*.
- Margolin, D. B., Hannak, A., and Weber, I. (2018). Political fact-checking on twitter: When do corrections have an effect? *Political Communication*, 35(2):196–219.
- Nyhan, B., Porter, E., Reifler, J., and Wood, T. J. (2020). Taking fact-checks literally but not seriously? the effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, 42:939–960.
- 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.
- Swire-Thompson, B., DeGutis, J., and Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of applied research in memory and cognition*, 9(3):286–299.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological review*, 109(3):451.

- 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.

Supplementary information

Contents

1	Treatment designs	2
2	Balance tests for all specifications	7
3	Differences of means in the general specification	21
4	Differences of means in secondary specifications	26
5	Heterogeneity in results	34
5.1	Partisan attachment and vaccination status	37
5.2	Education and attention	48
6	Placebo experiment	58

1 Treatment designs

As explained in the main document, respondents were exposed to a fictional Facebook post consisting of an image and a brief text. The text was either a confirmation of a correct statement or a refutation of an incorrect one. The main treatment communicates that vaccines are effective against Omicron in both the confirmation and refutation frames, while the placebo experiment reads that dogs do not understand human conversations. Variants of the main treatment consist in a rotation of the vaccine brand mentioned (being AstraZeneca, Moderna, and Sputnik V) in the Argentine experiment (see Figure S1) and the inclusion or not of a ‘TRUE’ or ‘FALSE’ label printed over the post image (a common practice in real fact-checking posts) in the Brazilian, Chilean, and Colombian experiments (see Figures S2 through S4). For details on the placebo experiment, refer to Section 6 of this document.

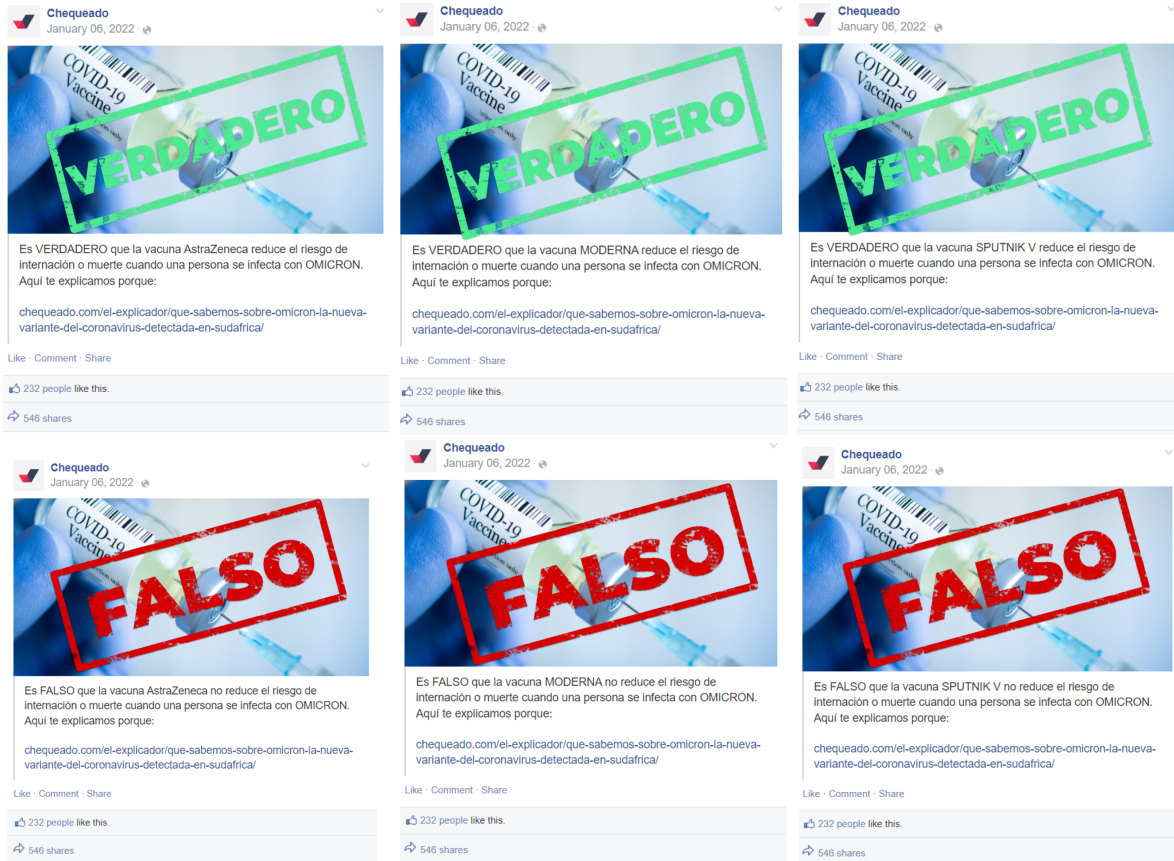


Figure S1 Images of the *confirmation* and *refutation* treatments used in Argentina. The confirmation and refutation frames are semantically equivalent but differ in their cognitive accessibility and their valence charge. The texts read “It is [TRUE/FALSE] that the [AstraZeneca/Moderna/Sputnik V] vaccine [reduces/does not reduce] the risk of hospitalization or death when a person gets infected with OMICRON”. All six treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*.



Figure S2 Images of the *confirmation* and *refutation* vaccine treatments used in Brazil. The confirmation and refutation frames are semantically equivalent but differ in their cognitive accessibility and their valence charge. The texts read “It is [TRUE/FALSE] that Pfizer-BioNTech’s new BivalentVaccine [is/is not] effective against the Omicron variant, contrary to what has been circulated on social media recently”. Treatments on the left do not include a printed label and treatments on the right do include them. All four treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*.



Figure S3 Images of the *confirmation* and *refutation* vaccine treatments used in Chile. The confirmation and refutation frames are semantically equivalent but differ in their cognitive accessibility and their valence charge. The texts read “It is [TRUE/FALSE] that Pfizer-BioNTech’s new BivalentVaccine [is/is not] effective against the Omicron variant, contrary to what has been circulated on social media recently”. Treatments on the left do not include a printed label and treatments on the right do include them. All four treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*.



Figure S4 Images of the *confirmation* and *refutation* vaccine treatments used in Colombia. The confirmation and refutation frames are semantically equivalent but differ in their cognitive accessibility and their valence charge. The texts read “It is [TRUE/FALSE] that Pfizer-BioNTech’s new BivalentVaccine [is/is not] effective against the Omicron variant, contrary to what has been circulated on social media recently”. Treatments on the left do not include a printed label and treatments on the right do include them. All four treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*.

2 Balance tests for all specifications

Tables [S1](#) to [S4](#) describe the differences in sample means for key socio-demographic and partisan indicators used in this SIF file. Differences between average values of each control variable are computed for the confirmation framing versus refutation framing groups for the general specification, the specifications with and without printed ‘TRUE’ and ‘FALSE’ labels, the specifications with different vaccine brands, and the placebo experiment, respectively. Except for age (measured in years) and the logarithm of time spent reading, all variables are indicators and the differences presented are in percentage points.

Table S1 Balance tests for control variables in the general specification by country. Difference of means of selected control variables by confirmation and refutation frame assignment.

Variable	Refutation frame	Confirmation frame	Difference
<i>Argentina</i>			
Woman	0.549 (0.015)	0.573 (0.014)	0.024 (0.020)
Age	40.664 (0.369)	40.294 (0.366)	-0.370 (0.519)
Incomplete secondary (or -)	0.063 (0.007)	0.052 (0.006)	-0.012 (0.010)
Completed secondary	0.190 (0.012)	0.203 (0.012)	0.013 (0.016)
Incomplete college	0.321 (0.014)	0.322 (0.014)	0.002 (0.019)
Completed college	0.350 (0.014)	0.359 (0.014)	0.009 (0.020)
Incomplete graduate (or +)	0.076 (0.008)	0.064 (0.007)	-0.011 (0.011)
Employed	0.799 (0.012)	0.801 (0.012)	0.001 (0.017)
Attachment to government	0.273 (0.013)	0.242 (0.012)	-0.031* (0.018)
Attachment to opposition	0.440 (0.015)	0.449 (0.014)	0.009 (0.021)
Log time to read	2.838 (0.025)	2.778 (0.022)	-0.060* (0.033)
Have had COVID-19	0.435 (0.015)	0.442 (0.014)	0.006 (0.020)
Non-vaccinated against COVID-19	0.058 (0.007)	0.062 (0.007)	0.003 (0.010)
Vaccinated once	0.029 (0.005)	0.033 (0.005)	0.005 (0.007)
Vaccinated twice (or +)	0.913 (0.008)	0.905 (0.009)	-0.008 (0.012)
<i>Brazil</i>			
Woman	0.496 (0.017)	0.476 (0.017)	-0.020 (0.025)
Age	39.440 (0.420)	39.962 (0.413)	0.522 (0.589)

Incomplete secondary (or -)	0.097 (0.010)	0.086 (0.010)	-0.011 (0.014)
Completed secondary	0.281 (0.016)	0.309 (0.016)	0.028 (0.022)
Incomplete college	0.143 (0.012)	0.127 (0.012)	-0.016 (0.017)
Completed college	0.264 (0.015)	0.275 (0.016)	0.011 (0.022)
Incomplete graduate (or +)	0.215 (0.014)	0.201 (0.014)	-0.013 (0.020)
Employed	0.772 (0.015)	0.774 (0.015)	0.002 (0.021)
Attachment to government	0.384 (0.017)	0.386 (0.017)	0.002 (0.024)
Attachment to opposition	0.421 (0.017)	0.431 (0.017)	0.010 (0.024)
Log time to read	2.795 (0.027)	2.824 (0.027)	0.029 (0.038)
Have had COVID-19	0.451 (0.017)	0.480 (0.017)	0.029 (0.025)
Non-vaccinated against COVID-19	0.057 (0.008)	0.049 (0.008)	-0.008 (0.011)
Vaccinated once	0.043 (0.007)	0.033 (0.006)	-0.010 (0.009)
Vaccinated twice (or +)	0.900 (0.010)	0.918 (0.010)	0.018 (0.014)

Chile

Woman	0.598 (0.017)	0.577 (0.017)	-0.021 (0.024)
Age	42.132 (0.434)	41.801 (0.436)	-0.331 (0.615)
Incomplete secondary (or -)	0.015 (0.004)	0.020 (0.005)	0.005 (0.006)
Completed secondary	0.205 (0.014)	0.209 (0.014)	0.004 (0.020)
Incomplete college	0.207 (0.014)	0.212 (0.014)	0.004 (0.020)
Completed college	0.466 (0.018)	0.442 (0.017)	-0.025 (0.025)
Incomplete graduate (or +)	0.107	0.118	0.011

	(0.011)	(0.011)	(0.016)
Employed	0.770	0.794	0.024
	(0.015)	(0.014)	(0.021)
Attachment to government	0.430	0.438	0.008
	(0.017)	(0.017)	(0.025)
Attachment to opposition	0.358	0.323	-0.034
	(0.017)	(0.016)	(0.024)
Log time to read	2.924	3.058	0.134***
	(0.024)	(0.028)	(0.036)
Have had COVID-19	0.399	0.403	0.004
	(0.017)	(0.017)	(0.024)
Non-vaccinated against COVID-19	0.036	0.041	0.005
	(0.007)	(0.007)	(0.010)
Vaccinated once	0.012	0.014	0.001
	(0.004)	(0.004)	(0.006)
Vaccinated twice (or +)	0.952	0.946	-0.006
	(0.008)	(0.008)	(0.011)
<hr/>			
<i>Colombia</i>			
Woman	0.522	0.549	0.028
	(0.018)	(0.018)	(0.025)
Age	37.814	37.424	-0.390
	(0.418)	(0.453)	(0.616)
Incomplete secondary (or -)	0.027	0.035	0.007
	(0.006)	(0.007)	(0.009)
Completed secondary	0.229	0.223	-0.006
	(0.015)	(0.015)	(0.021)
Incomplete college	0.236	0.236	-0.000
	(0.015)	(0.016)	(0.022)
Completed college	0.401	0.404	0.003
	(0.017)	(0.018)	(0.025)
Incomplete graduate (or +)	0.107	0.103	-0.004
	(0.011)	(0.011)	(0.016)
Employed	0.809	0.841	0.032*
	(0.014)	(0.013)	(0.019)
Attachment to government	0.512	0.472	-0.040
	(0.018)	(0.018)	(0.025)
Attachment to opposition	0.201	0.181	-0.020
	(0.014)	(0.014)	(0.020)
Log time to read	2.974	3.067	0.093**
	(0.026)	(0.028)	(0.038)

Have had COVID-19	0.413 (0.017)	0.427 (0.018)	0.014 (0.025)
Non-vaccinated against COVID-19	0.054 (0.008)	0.071 (0.009)	0.017 (0.012)
Vaccinated once	0.101 (0.011)	0.110 (0.011)	0.009 (0.016)
Vaccinated twice (or +)	0.846 (0.013)	0.819 (0.014)	-0.026 (0.019)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first and second columns represent the average value of each control variable for the refutation frame and confirmation frame groups respectively. Significance levels of differences presented in the third column are derived from linear regression models without controls.

Table S2 Balance tests for control variables in the Brazil, Chile, and Colombia surveys by use of labels. Difference of means of selected control variables by confirmation and refutation frame assignment for the specifications with and without printed ‘TRUE’ and ‘FALSE’ labels.

Variable	With labels			Without labels		
	Refutation	Confirmation	Difference	Refutation	Confirmation	Difference
<i>Brazil</i>						
Woman	0.511 (0.025)	0.486 (0.025)	-0.025 (0.035)	0.480 (0.025)	0.467 (0.024)	-0.014 (0.034)
Age	39.128 (0.595)	40.036 (0.590)	0.908 (0.838)	39.756 (0.593)	39.897 (0.578)	0.140 (0.828)
Incomplete secondary (or -)	0.106 (0.015)	0.093 (0.015)	-0.013 (0.021)	0.088 (0.014)	0.080 (0.013)	-0.007 (0.019)
Completed secondary	0.280 (0.022)	0.301 (0.023)	0.021 (0.032)	0.283 (0.022)	0.317 (0.022)	0.034 (0.032)
Incomplete college	0.147 (0.017)	0.129 (0.017)	-0.018 (0.024)	0.139 (0.017)	0.126 (0.016)	-0.013 (0.023)
Completed college	0.260 (0.022)	0.260 (0.022)	-0.001 (0.031)	0.268 (0.022)	0.290 (0.022)	0.021 (0.031)
Incomplete graduate (or +)	0.207 (0.020)	0.219 (0.021)	0.011 (0.029)	0.222 (0.021)	0.186 (0.019)	-0.036 (0.028)
Employed	0.780 (0.020)	0.768 (0.021)	-0.012 (0.030)	0.763 (0.021)	0.779 (0.020)	0.016 (0.029)
Attachment to government	0.381 (0.024)	0.414 (0.025)	0.033 (0.035)	0.388 (0.024)	0.361 (0.023)	-0.027 (0.033)
Attachment to opposition	0.417 (0.024)	0.398 (0.025)	-0.018 (0.035)	0.424 (0.024)	0.460 (0.024)	0.035 (0.034)
Log time to read	2.800 (0.038)	2.855 (0.039)	0.056 (0.054)	2.790 (0.037)	2.796 (0.037)	0.005 (0.052)
Have had COVID-19	0.448	0.488	0.041	0.454	0.472	0.018

Non-vaccinated against COVID-19	(0.025)	(0.025)	(0.035)	(0.025)	(0.024)	(0.034)
	0.056	0.070	0.014	0.059	0.030	-0.028**
	(0.011)	(0.013)	(0.017)	(0.012)	(0.008)	(0.014)
Vaccinated once	0.034	0.034	-0.000	0.051	0.032	-0.019
	(0.009)	(0.009)	(0.013)	(0.011)	(0.009)	(0.014)
Vaccinated twice (or +)	0.910	0.897	-0.014	0.890	0.938	0.047**
	(0.014)	(0.015)	(0.021)	(0.015)	(0.012)	(0.019)
<i>Chile</i>						
Woman	0.583	0.570	-0.013	0.612	0.584	-0.028
	(0.025)	(0.025)	(0.035)	(0.024)	(0.024)	(0.034)
Age	41.046	42.075	1.029	43.153	41.533	-1.620*
	(0.627)	(0.620)	(0.881)	(0.598)	(0.613)	(0.857)
Incomplete secondary (or -)	0.015	0.020	0.005	0.014	0.019	0.005
	(0.006)	(0.007)	(0.009)	(0.006)	(0.007)	(0.009)
Completed secondary	0.209	0.204	-0.005	0.201	0.214	0.013
	(0.021)	(0.020)	(0.029)	(0.020)	(0.020)	(0.028)
Incomplete college	0.216	0.201	-0.015	0.199	0.221	0.023
	(0.021)	(0.020)	(0.029)	(0.020)	(0.021)	(0.028)
Completed college	0.455	0.468	0.012	0.476	0.416	-0.060*
	(0.025)	(0.025)	(0.035)	(0.024)	(0.024)	(0.035)
Incomplete graduate (or +)	0.104	0.107	0.003	0.110	0.129	0.019
	(0.015)	(0.015)	(0.022)	(0.015)	(0.017)	(0.023)
Employed	0.762	0.823	0.061**	0.778	0.766	-0.011
	(0.022)	(0.019)	(0.029)	(0.020)	(0.021)	(0.029)
Attachment to government	0.420	0.438	0.018	0.440	0.438	-0.002
	(0.025)	(0.025)	(0.035)	(0.024)	(0.025)	(0.035)
Attachment to opposition	0.359	0.343	-0.015	0.356	0.304	-0.052
	(0.024)	(0.024)	(0.034)	(0.023)	(0.023)	(0.033)
Log time to read	2.868	3.087	0.220***	2.977	3.029	0.052

Have had COVID-19	(0.032)	(0.040)	(0.051)	(0.034)	(0.038)	(0.052)
	0.380	0.411	0.031	0.416	0.394	-0.022
	(0.025)	(0.025)	(0.035)	(0.024)	(0.024)	(0.034)
Non-vaccinated against COVID-19	0.043	0.042	-0.001	0.029	0.039	0.010
	(0.010)	(0.010)	(0.014)	(0.008)	(0.010)	(0.013)
Vaccinated once	0.013	0.012	-0.000	0.012	0.015	0.003
	(0.006)	(0.006)	(0.008)	(0.005)	(0.006)	(0.008)
Vaccinated twice (or +)	0.944	0.945	0.001	0.959	0.946	-0.013
	(0.012)	(0.011)	(0.016)	(0.010)	(0.011)	(0.015)

Colombia

Woman	0.481	0.525	0.044	0.561	0.573	0.012
	(0.025)	(0.026)	(0.036)	(0.025)	(0.026)	(0.035)
Age	38.342	38.008	-0.334	37.305	36.846	-0.459
	(0.608)	(0.665)	(0.901)	(0.574)	(0.614)	(0.840)
Incomplete secondary (or -)	0.025	0.040	0.015	0.029	0.029	-0.000
	(0.008)	(0.010)	(0.013)	(0.008)	(0.009)	(0.012)
Completed secondary	0.225	0.252	0.027	0.232	0.194	-0.038
	(0.021)	(0.023)	(0.031)	(0.021)	(0.020)	(0.029)
Incomplete college	0.225	0.214	-0.011	0.246	0.257	0.011
	(0.021)	(0.021)	(0.030)	(0.021)	(0.023)	(0.031)
Completed college	0.400	0.402	0.002	0.402	0.406	0.003
	(0.025)	(0.025)	(0.035)	(0.024)	(0.025)	(0.035)
Incomplete graduate (or +)	0.124	0.091	-0.033	0.090	0.114	0.024
	(0.017)	(0.015)	(0.022)	(0.014)	(0.016)	(0.022)
Employed	0.820	0.845	0.024	0.798	0.838	0.040
	(0.019)	(0.019)	(0.027)	(0.020)	(0.019)	(0.028)
Attachment to government	0.522	0.458	-0.063*	0.502	0.485	-0.017
	(0.025)	(0.026)	(0.036)	(0.025)	(0.026)	(0.036)
Attachment to opposition	0.220	0.174	-0.046	0.183	0.188	0.005

Log time to read	(0.021)	(0.020)	(0.029)	(0.019)	(0.020)	(0.028)
	2.984	3.091	0.107*	2.963	3.043	0.080
	(0.039)	(0.041)	(0.057)	(0.036)	(0.036)	(0.051)
Have had COVID-19	0.420	0.410	-0.010	0.407	0.444	0.037
	(0.025)	(0.026)	(0.036)	(0.024)	(0.026)	(0.035)
Non-vaccinated against COVID-19	0.063	0.073	0.009	0.044	0.069	0.025
	(0.012)	(0.014)	(0.018)	(0.010)	(0.013)	(0.017)
Vaccinated once	0.107	0.108	0.001	0.095	0.112	0.016
	(0.016)	(0.016)	(0.022)	(0.015)	(0.016)	(0.022)
Vaccinated twice (or +)	0.830	0.819	-0.011	0.861	0.819	-0.041
	(0.019)	(0.020)	(0.028)	(0.017)	(0.020)	(0.026)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns labeled as Refutation and Confirmation present the average value of each control variable for the refutation frame and confirmation frame groups respectively, either in the treatment branch that includes visual labels of TRUE and FALSE or in the treatment branch that does not include the labels. Significance levels of the differences presented are derived from linear regression models without controls.

Table S3 Balance tests for control variables in the Argentina survey by mentioned vaccine brand. Difference of means of selected control variables by confirmation and refutation frame assignment for each brand of vaccines mentioned in the vignette in the Argentina survey.

Variable	Refutation frame	Confirmation frame	Difference
<i>Moderna</i>			
Woman	0.544 (0.025)	0.569 (0.025)	0.025 (0.035)
Age	40.894 (0.656)	40.180 (0.631)	-0.714 (0.910)
Incomplete secondary (or -)	0.065 (0.012)	0.058 (0.012)	-0.008 (0.017)
Completed secondary	0.191 (0.020)	0.198 (0.020)	0.007 (0.028)
Incomplete college	0.325 (0.024)	0.318 (0.023)	-0.007 (0.033)
Completed college	0.343 (0.024)	0.363 (0.024)	0.021 (0.034)
Incomplete graduate (or +)	0.076 (0.013)	0.063 (0.012)	-0.013 (0.018)
Employed	0.778 (0.021)	0.817 (0.019)	0.039 (0.029)
Attachment to government	0.267 (0.022)	0.228 (0.021)	-0.039 (0.031)
Attachment to opposition	0.463 (0.025)	0.464 (0.025)	0.000 (0.035)
Log time to read	2.867 (0.043)	2.794 (0.040)	-0.074 (0.059)
Have had COVID-19	0.441 (0.025)	0.446 (0.025)	0.005 (0.035)
Non-vaccinated against COVID-19	0.048 (0.011)	0.065 (0.012)	0.018 (0.016)
Vaccinated once	0.028 (0.008)	0.038 (0.010)	0.010 (0.013)
Vaccinated twice (or +)	0.924 (0.013)	0.897 (0.015)	-0.028 (0.020)
<i>Sputnik</i>			
Woman	0.540 (0.027)	0.558 (0.025)	0.018 (0.036)
Age	40.195	40.276	0.081

	(0.658)	(0.616)	(0.901)
Incomplete secondary (or -)	0.065	0.045	-0.020
	(0.013)	(0.010)	(0.017)
Completed secondary	0.158	0.221	0.063**
	(0.019)	(0.021)	(0.028)
Incomplete college	0.342	0.314	-0.028
	(0.025)	(0.023)	(0.034)
Completed college	0.364	0.364	-0.000
	(0.026)	(0.024)	(0.035)
Incomplete graduate (or +)	0.071	0.055	-0.015
	(0.014)	(0.011)	(0.018)
Employed	0.807	0.789	-0.018
	(0.021)	(0.020)	(0.029)
Attachment to government	0.268	0.264	-0.005
	(0.024)	(0.022)	(0.032)
Attachment to opposition	0.460	0.417	-0.043
	(0.027)	(0.025)	(0.036)
Log time to read	2.781	2.736	-0.045
	(0.044)	(0.036)	(0.057)
Have had COVID-19	0.455	0.435	-0.020
	(0.027)	(0.025)	(0.036)
Non-vaccinated against COVID-19	0.074	0.060	-0.014
	(0.014)	(0.012)	(0.018)
Vaccinated once	0.026	0.038	0.012
	(0.008)	(0.010)	(0.013)
Vaccinated twice (or +)	0.900	0.902	0.002
	(0.016)	(0.015)	(0.022)
<hr/>			
<i>AstraZeneca</i>			
Woman	0.563	0.591	0.029
	(0.025)	(0.025)	(0.035)
Age	40.850	40.424	-0.426
	(0.602)	(0.656)	(0.891)
Incomplete secondary (or -)	0.060	0.052	-0.008
	(0.012)	(0.011)	(0.016)
Completed secondary	0.217	0.190	-0.028
	(0.021)	(0.020)	(0.028)
Incomplete college	0.297	0.334	0.037
	(0.023)	(0.024)	(0.033)
Completed college	0.345	0.349	0.004
	(0.024)	(0.024)	(0.034)

Incomplete graduate (or +)	0.080 (0.014)	0.075 (0.013)	-0.005 (0.019)
Employed	0.814 (0.020)	0.796 (0.020)	-0.019 (0.028)
Attachment to government	0.282 (0.023)	0.234 (0.021)	-0.048 (0.031)
Attachment to opposition	0.398 (0.024)	0.466 (0.025)	0.069** (0.035)
Log time to read	2.860 (0.041)	2.805 (0.036)	-0.055 (0.054)
Have had COVID-19	0.412 (0.025)	0.444 (0.025)	0.031 (0.035)
Non-vaccinated against COVID-19	0.055 (0.011)	0.060 (0.012)	0.005 (0.016)
Vaccinated once	0.033 (0.009)	0.025 (0.008)	-0.008 (0.012)
Vaccinated twice (or +)	0.912 (0.014)	0.915 (0.014)	0.003 (0.020)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first and second columns represent the average value of each control variable for the refutation frame and confirmation frame groups respectively. Significance levels of differences presented in the third column are derived from linear regression models without controls.

Table S4 Balance tests for control variables in the placebo experiment by country. Difference of means of selected control variables by confirmation and refutation frame assignment.

Variable	Refutation frame	Confirmation frame	Difference
<i>Brazil</i>			
Woman	0.501 (0.026)	0.530 (0.025)	0.029 (0.036)
Age	39.295 (0.626)	39.519 (0.633)	0.224 (0.890)
Incomplete secondary (or -)	0.074 (0.014)	0.104 (0.016)	0.030 (0.021)
Completed secondary	0.259 (0.023)	0.268 (0.023)	0.009 (0.032)
Incomplete college	0.185 (0.020)	0.166 (0.019)	-0.019 (0.028)
Completed college	0.300	0.268	-0.032

	(0.024)	(0.023)	(0.033)
Incomplete graduate (or +)	0.183	0.195	0.012
	(0.020)	(0.020)	(0.029)
Employed	0.747	0.706	-0.041
	(0.023)	(0.023)	(0.033)
Attachment to government	0.401	0.348	-0.052
	(0.026)	(0.024)	(0.035)
Attachment to opposition	0.403	0.455	0.051
	(0.026)	(0.025)	(0.036)
Log time to read	2.690	2.693	0.003
	(0.036)	(0.035)	(0.050)
Have had COVID-19	0.417	0.413	-0.004
	(0.026)	(0.025)	(0.036)
Non-vaccinated against COVID-19	0.052	0.044	-0.008
	(0.012)	(0.010)	(0.016)
Vaccinated once	0.027	0.036	0.009
	(0.009)	(0.010)	(0.013)
Vaccinated twice (or +)	0.921	0.919	-0.002
	(0.014)	(0.014)	(0.020)
<hr/>			
<i>Chile</i>			
Woman	0.557	0.554	-0.003
	(0.026)	(0.025)	(0.036)
Age	43.095	41.337	-1.758*
	(0.663)	(0.622)	(0.909)
Incomplete secondary (or -)	0.014	0.012	-0.001
	(0.006)	(0.006)	(0.008)
Completed secondary	0.142	0.202	0.060**
	(0.018)	(0.020)	(0.027)
Incomplete college	0.240	0.217	-0.023
	(0.023)	(0.021)	(0.031)
Completed college	0.496	0.476	-0.020
	(0.026)	(0.025)	(0.036)
Incomplete graduate (or +)	0.109	0.092	-0.016
	(0.016)	(0.014)	(0.022)
Employed	0.816	0.812	-0.004
	(0.020)	(0.020)	(0.028)
Attachment to government	0.398	0.394	-0.004
	(0.026)	(0.024)	(0.036)
Attachment to opposition	0.370	0.357	-0.014
	(0.026)	(0.024)	(0.035)

Log time to read	2.972 (0.038)	2.861 (0.036)	-0.111** (0.052)
Have had COVID-19	0.394 (0.026)	0.406 (0.025)	0.013 (0.036)
Non-vaccinated against COVID-19	0.061 (0.013)	0.057 (0.012)	-0.004 (0.017)
Vaccinated once	0.014 (0.006)	0.012 (0.006)	-0.001 (0.008)
Vaccinated twice (or +)	0.925 (0.014)	0.930 (0.013)	0.006 (0.019)
<hr/>			
<i>Colombia</i>			
Woman	0.531 (0.025)	0.548 (0.024)	0.017 (0.035)
Age	37.174 (0.600)	37.632 (0.595)	0.458 (0.845)
Incomplete secondary (or -)	0.027 (0.008)	0.029 (0.008)	0.002 (0.012)
Completed secondary	0.184 (0.019)	0.200 (0.020)	0.015 (0.027)
Incomplete college	0.253 (0.022)	0.228 (0.021)	-0.025 (0.030)
Completed college	0.418 (0.024)	0.413 (0.024)	-0.004 (0.034)
Incomplete graduate (or +)	0.118 (0.016)	0.130 (0.016)	0.012 (0.023)
Employed	0.811 (0.019)	0.827 (0.019)	0.016 (0.027)
Attachment to government	0.491 (0.025)	0.474 (0.025)	-0.018 (0.035)
Attachment to opposition	0.182 (0.019)	0.151 (0.018)	-0.030 (0.026)
Log time to read	2.864 (0.035)	2.808 (0.033)	-0.056 (0.049)
Have had COVID-19	0.424 (0.025)	0.463 (0.025)	0.039 (0.035)
Non-vaccinated against COVID-19	0.067 (0.012)	0.056 (0.011)	-0.011 (0.017)
Vaccinated once	0.086 (0.014)	0.092 (0.014)	0.006 (0.020)
Vaccinated twice (or +)	0.847 (0.018)	0.853 (0.017)	0.005 (0.025)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first and second columns represent the average value of each control variable for the refutation frame and confirmation frame groups respectively in the placebo experiment. Significance levels of differences presented in the third column are derived from linear regression models without controls.

3 Differences of means in the general specification

Tables S5 to S8 expand the results from Table 1 of the main article. They present the detailed differences of means of the dependent variables (reactions and emotions generated) between the confirmation and refutation frames in the general specification for each specific country survey (Argentina, Brazil, Chile, and Colombia, respectively). Both frames are semantically equivalent propositions phrased as a confirmation of accurate information or as refutation of inaccurate information. The results in these tables can be used to compare findings in the complementary specifications (varying the vaccine brand, the use of printed ‘TRUE’ or ‘FALSE’ labels, heterogeneous specifications, and the placebo experiment).

Table S5 Difference of means between the confirmation and refutation frames in the Argentina survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.189 (0.012)	0.371 (0.014)	0.182*** (0.018)	0.188*** (0.018)
Like	0.084 (0.008)	0.245 (0.012)	0.160*** (0.015)	0.163*** (0.015)
Share	0.076 (0.008)	0.115 (0.009)	0.039*** (0.012)	0.042*** (0.012)
Comment	0.043 (0.006)	0.048 (0.006)	0.006 (0.009)	0.008 (0.009)
<i>Emotions</i>				
Optimistic	0.060 (0.007)	0.333 (0.014)	0.273*** (0.015)	0.278*** (0.015)
Joyful	0.017 (0.004)	0.098 (0.009)	0.082*** (0.009)	0.087*** (0.009)
Angry	0.165 (0.011)	0.045 (0.006)	-0.120*** (0.012)	-0.121*** (0.012)
Sad	0.058 (0.007)	0.023 (0.004)	-0.035*** (0.008)	-0.035*** (0.008)
Stressed	0.128 (0.010)	0.058 (0.007)	-0.069*** (0.012)	-0.072*** (0.012)
Fearful	0.043 (0.006)	0.028 (0.005)	-0.015* (0.008)	-0.015* (0.008)
Disgusted	0.195 (0.012)	0.076 (0.008)	-0.119*** (0.014)	-0.121*** (0.014)
Indifferent	0.443 (0.015)	0.417 (0.014)	-0.027 (0.020)	-0.028 (0.020)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

Table S6 Difference of means between the confirmation and refutation frames in the Brazil survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.412 (0.017)	0.539 (0.017)	0.127*** (0.024)	0.131*** (0.024)
Like	0.275 (0.016)	0.402 (0.017)	0.127*** (0.023)	0.127*** (0.022)
Share	0.148 (0.012)	0.147 (0.012)	-0.001 (0.017)	0.001 (0.018)
Comment	0.104 (0.011)	0.104 (0.011)	0.000 (0.015)	-0.002 (0.015)
<i>Emotions</i>				
Optimistic	0.178 (0.013)	0.444 (0.017)	0.266*** (0.022)	0.258*** (0.022)
Joyful	0.103 (0.011)	0.210 (0.014)	0.107*** (0.018)	0.106*** (0.018)
Angry	0.125 (0.012)	0.040 (0.007)	-0.085*** (0.013)	-0.083*** (0.014)
Sad	0.136 (0.012)	0.033 (0.006)	-0.103*** (0.013)	-0.102*** (0.014)
Stressed	0.091 (0.010)	0.036 (0.007)	-0.055*** (0.012)	-0.051*** (0.012)
Fearful	0.115 (0.011)	0.059 (0.008)	-0.056*** (0.014)	-0.051*** (0.014)
Disgusted	0.147 (0.012)	0.038 (0.007)	-0.109*** (0.014)	-0.112*** (0.014)
Indifferent	0.343 (0.017)	0.316 (0.016)	-0.027 (0.023)	-0.029 (0.023)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

Table S7 Difference of means between the confirmation and refutation frames in the Chile survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.219 (0.015)	0.373 (0.017)	0.153*** (0.022)	0.152*** (0.023)
Like	0.086 (0.010)	0.257 (0.015)	0.171*** (0.018)	0.171*** (0.018)
Share	0.122 (0.012)	0.137 (0.012)	0.014 (0.017)	0.010 (0.017)
Comment	0.039 (0.007)	0.048 (0.007)	0.009 (0.010)	0.006 (0.010)
<i>Emotions</i>				
Optimistic	0.095 (0.010)	0.338 (0.017)	0.243*** (0.020)	0.246*** (0.020)
Joyful	0.028 (0.006)	0.135 (0.012)	0.107*** (0.013)	0.105*** (0.014)
Angry	0.144 (0.012)	0.062 (0.008)	-0.083*** (0.015)	-0.081*** (0.015)
Sad	0.043 (0.007)	0.018 (0.005)	-0.025*** (0.009)	-0.023*** (0.009)
Stressed	0.113 (0.011)	0.086 (0.010)	-0.027* (0.015)	-0.025* (0.015)
Fearful	0.116 (0.011)	0.074 (0.009)	-0.042*** (0.015)	-0.042*** (0.015)
Disgusted	0.137 (0.012)	0.091 (0.010)	-0.046*** (0.016)	-0.046*** (0.015)
Indifferent	0.503 (0.018)	0.406 (0.017)	-0.097*** (0.025)	-0.102*** (0.025)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

Table S8 Difference of means between the confirmation and refutation frames in the Colombia survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.289 (0.016)	0.428 (0.018)	0.139*** (0.024)	0.147*** (0.024)
Like	0.124 (0.012)	0.240 (0.016)	0.116*** (0.019)	0.120*** (0.020)
Share	0.152 (0.013)	0.184 (0.014)	0.032* (0.019)	0.035* (0.019)
Comment	0.051 (0.008)	0.077 (0.010)	0.026** (0.012)	0.026** (0.012)
<i>Emotions</i>				
Optimistic	0.142 (0.012)	0.363 (0.018)	0.221*** (0.021)	0.226*** (0.022)
Joyful	0.032 (0.006)	0.101 (0.011)	0.069*** (0.013)	0.071*** (0.013)
Angry	0.096 (0.010)	0.047 (0.008)	-0.049*** (0.013)	-0.049*** (0.013)
Sad	0.078 (0.009)	0.032 (0.006)	-0.046*** (0.011)	-0.049*** (0.011)
Stressed	0.108 (0.011)	0.080 (0.010)	-0.028* (0.015)	-0.028* (0.015)
Fearful	0.173 (0.013)	0.140 (0.013)	-0.033* (0.018)	-0.037** (0.019)
Disgusted	0.051 (0.008)	0.051 (0.008)	-0.000 (0.011)	0.001 (0.011)
Indifferent	0.516 (0.018)	0.381 (0.018)	-0.134*** (0.025)	-0.133*** (0.025)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

4 Differences of means in secondary specifications

Tables S9 to S14 expand on the results of the previous section with more detailed specifications. Tables S9 to S11 disaggregate confirmation framing effects by changing the vaccine brand mentioned in the fact-checking post. While, in general, results appear robust to the vaccine brand presented (Table S9), some differences emerge when they are broken down by partisan attachment (Tables S10 and S11). Consistent with the *Hypothesis 3* in the pre-approved plan, which described the independent pro- and counter-attitudinal effect of partisan attachment by type of vaccine, these results show a considerably higher relative effect for Moderna over Sputnik V and for AstraZeneca over Sputnik V among respondents who support the opposition, since in Argentina, the Sputnik V vaccine was typically associated with the government in the media. We expected that respondents who supported the government (i.e., *Frente de Todos (FdT)* voters) would be more inclined to share confirmations of Sputnik V than Moderna and AstraZeneca. We also expected that the opposition (i.e., voters of *Cambiamos*) would be more likely to share confirmation of Moderna than Sputnik and AstraZeneca. These effects have no associated cognitive cost. Differences in the overall effects on reactions between supporters of the government and the opposition are further detailed in section 5.1 of this Supplemental Information File.

Aside from the Argentina survey, the ones conducted in Brazil, Chile, and Colombia included different treatments by the use of printed ‘TRUE’ or ‘FALSE’ labels in the accompanying image. Tables S12 to S14 report differences in results by use of labels in each of the three surveys, extending the results from Table 2 in the main article. While the effects on emotions are stable to the inclusion or not of the label, engagement appears to be more volatile, increasing with the use of labels in the Brazil sample, decreasing in the Chile sample, and remaining about the same in the Colombia sample.

Table S9 Difference of means between the confirmation and refutation frames in the Argentina survey by brand of vaccine mentioned in the statement.

Variable	Overall	Moderna	Sputnik	AstraZeneca
<i>Reactions</i>				
Engage	0.188*** (0.018)	0.202*** (0.032)	0.177*** (0.032)	0.188*** (0.031)
Like	0.163*** (0.015)	0.174*** (0.027)	0.157*** (0.026)	0.163*** (0.025)
Share	0.042*** (0.012)	0.032 (0.020)	0.032 (0.023)	0.062*** (0.021)
Comment	0.008 (0.009)	0.019 (0.014)	0.023 (0.015)	-0.020 (0.016)
<i>Emotions</i>				
Optimistic	0.278*** (0.015)	0.250*** (0.027)	0.296*** (0.027)	0.296*** (0.026)
Joyful	0.087*** (0.009)	0.099*** (0.017)	0.107*** (0.018)	0.057*** (0.014)
Angry	-0.121*** (0.012)	-0.115*** (0.021)	-0.086*** (0.021)	-0.152*** (0.022)
Sad	-0.035*** (0.008)	-0.047*** (0.015)	-0.030* (0.015)	-0.025* (0.013)
Stressed	-0.072*** (0.012)	-0.070*** (0.020)	-0.083*** (0.021)	-0.066*** (0.022)
Fearful	-0.015* (0.008)	-0.008 (0.014)	-0.018 (0.013)	-0.021 (0.015)
Disgusted	-0.121*** (0.014)	-0.157*** (0.023)	-0.120*** (0.025)	-0.094*** (0.024)
Indifferent	-0.028 (0.020)	-0.009 (0.035)	-0.052 (0.036)	-0.029 (0.036)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction or emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S10 Difference of means in reactions between the confirmation and refutation frames in the Argentina survey by brand of vaccine mentioned in the statement and partisan attachment of the respondent.

Variable	Overall	Moderna	Sputnik	AstraZeneca
<i>Government (Frente de Todos)</i>				
Engage	0.269*** (0.037)	0.363*** (0.067)	0.310*** (0.066)	0.200*** (0.066)
Like	0.238*** (0.034)	0.315*** (0.064)	0.222*** (0.062)	0.241*** (0.060)
Share	0.057** (0.027)	0.069 (0.042)	0.094* (0.052)	0.005 (0.045)
Comment	0.024 (0.017)	0.011 (0.027)	0.072** (0.031)	-0.014 (0.033)
<i>Opposition (Juntos por el Cambio)</i>				
Engage	0.160*** (0.027)	0.205*** (0.046)	0.072 (0.050)	0.198*** (0.048)
Like	0.135*** (0.021)	0.166*** (0.039)	0.103*** (0.037)	0.146*** (0.037)
Share	0.042** (0.018)	0.048* (0.027)	-0.017 (0.035)	0.079** (0.031)
Comment	0.004 (0.014)	0.016 (0.022)	0.007 (0.023)	-0.005 (0.027)
<i>None (blank vote)</i>				
Engage	0.163*** (0.032)	0.097* (0.059)	0.240*** (0.059)	0.160*** (0.054)
Like	0.141*** (0.026)	0.100** (0.049)	0.200*** (0.050)	0.131*** (0.042)
Share	0.032 (0.021)	-0.023 (0.041)	0.064* (0.038)	0.068** (0.034)
Comment	0.001 (0.013)	0.021 (0.019)	0.016 (0.021)	-0.040* (0.024)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S11 Difference of means in emotions generation between the confirmation and refutation frames in the Argentina survey by brand of vaccine mentioned in the statement and partisan attachment of the respondent.

Variable	Overall	Moderna	Sputnik	AstraZeneca
<i>Government (Frente de Todos)</i>				
Optimistic	0.377*** (0.034)	0.361*** (0.064)	0.404*** (0.063)	0.404*** (0.059)
Joyful	0.118*** (0.023)	0.128*** (0.045)	0.092** (0.042)	0.125*** (0.040)
Angry	-0.177*** (0.027)	-0.204*** (0.054)	-0.150*** (0.046)	-0.190*** (0.046)
Sad	-0.052*** (0.017)	-0.091** (0.040)	-0.044 (0.027)	-0.027* (0.016)
Stressed	-0.102*** (0.023)	-0.108*** (0.041)	-0.100*** (0.037)	-0.123*** (0.045)
Fearful	-0.018* (0.011)	-0.011 (0.023)	-0.019 (0.014)	-0.030 (0.023)
Disgusted	-0.159*** (0.025)	-0.151*** (0.043)	-0.174*** (0.045)	-0.158*** (0.046)
Indifferent	0.005 (0.039)	0.004 (0.068)	0.010 (0.068)	0.001 (0.066)
<i>Opposition (Juntos por el Cambio)</i>				
Optimistic	0.266*** (0.021)	0.296*** (0.037)	0.266*** (0.039)	0.249*** (0.038)
Joyful	0.077*** (0.013)	0.076*** (0.022)	0.100*** (0.025)	0.062*** (0.022)
Angry	-0.115*** (0.019)	-0.102*** (0.033)	-0.074** (0.034)	-0.168*** (0.035)
Sad	-0.026** (0.012)	-0.047** (0.021)	-0.003 (0.022)	-0.016 (0.019)
Stressed	-0.094*** (0.018)	-0.087*** (0.028)	-0.111*** (0.031)	-0.080** (0.034)
Fearful	-0.011 (0.012)	0.004 (0.019)	-0.017 (0.022)	-0.028 (0.023)
Disgusted	-0.118*** (0.022)	-0.185*** (0.036)	-0.092** (0.039)	-0.092** (0.039)
Indifferent	-0.014 (0.031)	0.009 (0.053)	-0.044 (0.056)	-0.024 (0.056)

None (blank vote)

Optimistic	0.213*** (0.027)	0.097** (0.049)	0.282*** (0.045)	0.256*** (0.045)
Joyful	0.066*** (0.015)	0.112*** (0.031)	0.089*** (0.033)	-0.013 (0.014)
Angry	-0.086*** (0.020)	-0.075** (0.033)	-0.042 (0.033)	-0.106*** (0.040)
Sad	-0.033** (0.016)	-0.023 (0.026)	-0.038 (0.029)	-0.020 (0.028)
Stressed	-0.017 (0.022)	-0.006 (0.035)	-0.038 (0.045)	-0.008 (0.039)
Fearful	-0.017 (0.017)	-0.028 (0.031)	-0.005 (0.031)	-0.004 (0.030)
Disgusted	-0.095*** (0.026)	-0.118*** (0.044)	-0.128** (0.053)	-0.070 (0.046)
Indifferent	-0.061 (0.038)	-0.033 (0.069)	-0.115 (0.070)	-0.026 (0.067)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S12 Difference of means between the confirmation and refutation frames in the Brazil survey by inclusion or not of ‘TRUE’/‘FALSE’ label in the picture.

Variable	Overall	With label	Without label
<i>Reactions</i>			
Engage	0.110*** (0.020)	0.163*** (0.034)	0.089*** (0.034)
Like	0.108*** (0.018)	0.164*** (0.032)	0.088*** (0.032)
Share	0.000 (0.014)	-0.014 (0.025)	0.002 (0.025)
Comment	-0.003 (0.013)	0.009 (0.020)	-0.015 (0.022)
<i>Emotions</i>			
Optimistic	0.172*** (0.017)	0.274*** (0.030)	0.230*** (0.031)
Joyful	0.066*** (0.014)	0.127*** (0.025)	0.087*** (0.025)
Angry	-0.062*** (0.011)	-0.079*** (0.020)	-0.081*** (0.018)
Sad	-0.078*** (0.012)	-0.133*** (0.021)	-0.078*** (0.018)
Stressed	-0.034*** (0.009)	-0.061*** (0.017)	-0.036** (0.017)
Fearful	-0.042*** (0.010)	-0.060*** (0.020)	-0.039** (0.019)
Disgusted	-0.073*** (0.012)	-0.119*** (0.021)	-0.107*** (0.020)
Indifferent	-0.010 (0.020)	-0.031 (0.033)	-0.026 (0.032)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction or emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S13 Difference of means between the confirmation and refutation frames in the Chile survey by inclusion or not of ‘TRUE’/‘FALSE’ label in the picture.

Variable	Overall	With label	Without label
<i>Reactions</i>			
Engage	0.133*** (0.019)	0.104*** (0.032)	0.199*** (0.032)
Like	0.140*** (0.016)	0.158*** (0.025)	0.185*** (0.027)
Share	0.014 (0.014)	-0.035 (0.024)	0.056** (0.024)
Comment	0.007 (0.011)	0.003 (0.015)	0.004 (0.014)
<i>Emotions</i>			
Optimistic	0.178*** (0.015)	0.228*** (0.027)	0.270*** (0.029)
Joyful	0.075*** (0.011)	0.078*** (0.016)	0.132*** (0.021)
Angry	-0.063*** (0.012)	-0.102*** (0.023)	-0.063*** (0.020)
Sad	-0.008 (0.010)	-0.021* (0.012)	-0.028** (0.013)
Stressed	-0.022** (0.011)	-0.027 (0.023)	-0.022 (0.020)
Fearful	-0.032*** (0.011)	-0.038* (0.021)	-0.046** (0.021)
Disgusted	-0.034*** (0.011)	-0.039 (0.024)	-0.050** (0.020)
Indifferent	-0.075*** (0.020)	-0.065* (0.036)	-0.140*** (0.034)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction or emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S14 Difference of means between the confirmation and refutation frames in the Colombia survey by inclusion or not of ‘TRUE’/‘FALSE’ label in the picture.

Variable	Overall	With label	Without label
<i>Reactions</i>			
Engage	0.129*** (0.020)	0.153*** (0.034)	0.138*** (0.034)
Like	0.088*** (0.017)	0.111*** (0.027)	0.127*** (0.029)
Share	0.046*** (0.015)	0.037 (0.027)	0.035 (0.027)
Comment	0.029** (0.013)	0.034* (0.018)	0.015 (0.018)
<i>Emotions</i>			
Optimistic	0.159*** (0.017)	0.243*** (0.030)	0.198*** (0.031)
Joyful	0.054*** (0.011)	0.057*** (0.018)	0.082*** (0.019)
Angry	-0.015 (0.011)	-0.063*** (0.019)	-0.035** (0.017)
Sad	-0.030*** (0.012)	-0.053*** (0.017)	-0.042*** (0.015)
Stressed	-0.020* (0.010)	-0.034 (0.022)	-0.017 (0.020)
Fearful	-0.021 (0.013)	-0.009 (0.027)	-0.063** (0.025)
Disgusted	-0.003 (0.008)	-0.015 (0.017)	0.019 (0.014)
Indifferent	-0.103*** (0.020)	-0.135*** (0.036)	-0.125*** (0.035)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using as sample the branch of the treatment indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction or emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

5 Heterogeneity in results

Tables [S15](#) to [S21](#) present several disaggregations of the confirmation framing effect across political preferences, education, and attention, with the aim of exploring hypotheses about how this effect may or not depend on the issue salience, prior beliefs, or cognitive mechanisms.

Table [S15](#) is a continuation of the Table 3 in the main article, as it summarizes some confirmation framing effects on emotions generation across partisan attachment, educational attainment and time spent reading the question, by survey. More detailed information on the heterogeneity of effects in these dimensions is presented in Tables [S16](#) to [S27](#). In addition to the target variable for potential heterogeneity, all estimates are obtained controlling for socio-demographic variables (gender, age, educational attainment, and employment status), partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the question.

Table S15 Summary of heterogeneous effects on emotions generation in the general specification by country and selected dimensions. Difference of means of emotions generation by confirmation and refutation frame assignment.

Variable	Partisan attachment		Completed		Incomplete		Time to read	
	Government	Opposition	secondary (or -)	college (or +)	First quintile	Fifth quintile		
<i>Argentina</i>								
Optimistic	0.377***	0.266***	0.252***	0.288***	0.206***	0.372***		
Joyful	0.118***	0.077***	0.084***	0.088***	0.082***	0.062***		
Angry	-0.177***	-0.115***	-0.091***	-0.131***	-0.112***	-0.121***		
Sad	-0.052***	-0.026**	-0.041**	-0.033***	-0.022	-0.030		
Stressed	-0.102***	-0.094***	-0.052*	-0.080***	-0.056**	-0.090***		
Fearful	-0.018*	-0.011	-0.015	-0.014*	-0.019	-0.017		
Disgusted	-0.159***	-0.118***	-0.084***	-0.136***	-0.124***	-0.155***		
Indifferent	0.005	-0.014	-0.042	-0.022	0.016	-0.061		
<i>Brazil</i>								
Optimistic	0.174***	0.343***	0.165***	0.316***	0.194***	0.230***		
Joyful	0.071***	0.161***	0.066**	0.132***	0.069*	0.133***		
Angry	-0.043*	-0.127***	-0.071***	-0.093***	-0.082***	-0.079***		
Sad	-0.058***	-0.157***	-0.101***	-0.106***	-0.096***	-0.108***		
Stressed	-0.033	-0.093***	-0.027	-0.065***	-0.094***	-0.016		
Fearful	-0.029	-0.072***	-0.027	-0.066***	-0.055*	-0.056		
Disgusted	-0.034	-0.185***	-0.081***	-0.132***	-0.135***	-0.136***		
Indifferent	-0.067*	0.005	-0.005	-0.042	0.037	-0.006		
<i>Chile</i>								
Optimistic	0.245***	0.303***	0.159***	0.270***	0.257***	0.247***		
Joyful	0.100***	0.139***	0.068***	0.115***	0.137***	0.150***		
Angry	-0.096***	-0.069***	-0.083**	-0.083***	-0.066**	-0.106***		
Sad	-0.013	-0.026*	-0.023	-0.024**	-0.019	-0.028		
Stressed	-0.037	-0.033	-0.016	-0.027	-0.085**	0.014		
Fearful	-0.012	-0.075***	-0.073*	-0.036**	-0.059**	-0.025		
Disgusted	-0.069***	-0.028	0.021	-0.068***	-0.051	-0.030		

Indifferent	-0.092**	-0.129***	-0.010	-0.120***	-0.115*	-0.134**
<i>Colombia</i>						
Optimistic	0.245***	0.209***	0.237***	0.220***	0.210***	0.178***
Joyful	0.088***	0.085**	0.070**	0.069***	0.052	0.067**
Angry	-0.063***	-0.042	-0.032	-0.056***	-0.045	-0.045
Sad	-0.061***	-0.057**	-0.037	-0.053***	0.000	-0.048
Stressed	-0.048**	0.016	0.021	-0.049***	0.042	-0.029
Fearful	-0.012	-0.084*	-0.027	-0.040*	-0.026	-0.026
Disgusted	-0.018	0.008	0.043*	-0.019*	0.002	0.040*
Indifferent	-0.141***	-0.064	-0.207***	-0.101***	-0.186***	-0.151***

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

5.1 Partisan attachment and vaccination status

Tables S16 through S21 provide estimates of heterogeneous effects by political preferences and vaccination status. These results show a slight pattern of more pronounced effects for those who were expected to be more supportive of vaccination. For instance, a higher preference was expected from voters of the incumbent administration of Alberto Fernandez in Argentina, who actively supported quarantine measures early in the pandemic 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 pandemic. Both the leading opposition coalition voters (*Cambiamos*) and independent voters (blank vote) shared the confirmation frames at higher rates than the refutation frames, but the magnitude of the effect was significantly surpassed in the sample of government supporters (Tables S16 and S18). In the Argentine election survey, 1,080 respondents indicated a preference for *Cambiamos*, 613 voters indicated a preference for the *Frente de Todos*, and 727 respondents indicated a preference for blank voting (see Figure S5). These results are complementary to those presented in section 4 on differential effects by vaccine brand and partisan attachment in the Argentina survey. In line with this case, the Brazil survey was conducted during the presidency of Jair Bolsonaro, who openly opposed COVID-19 vaccination. As expected, his supporters show less sensitivity to the framing effect on engagement, although this difference is not statistically significant (Table S16).

The results presented in Tables S20 and S21 also show that the confirmation frame induces higher engagement and a higher rate of likes among respondents with higher vaccine doses. A caveat here is that people with two or more doses represent about 85% to 95% of the survey sample for each country, with Colombia being the one with the fewest respondents vaccinated twice or more times. This is consistent with a valence interpretation of the frame effect and inconsistent with the cognitive burden hypothesis, as discussed in the main article. Results show that the effect of the confirmation frame is more strongly associated with more positive emotions and less negative ones among respondents with 2 or more COVID-19 doses.

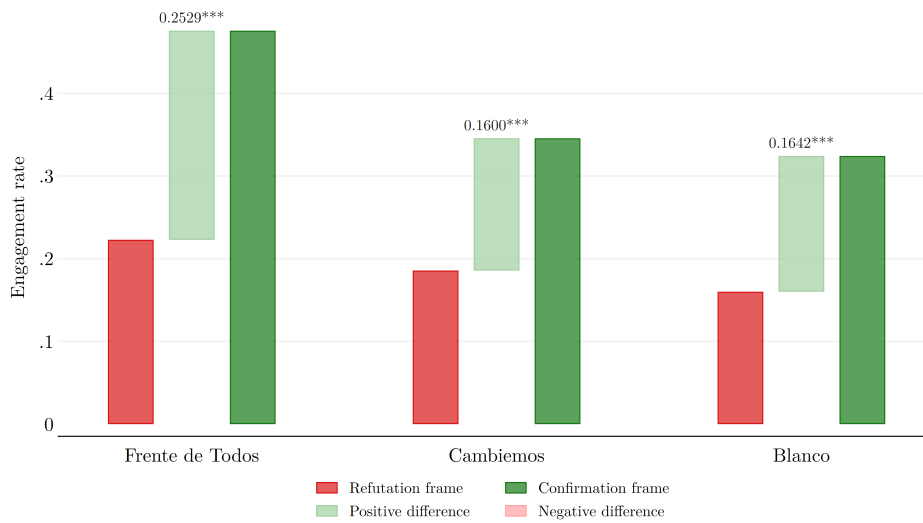


Figure S5 Engagement rates for the confirmation (‘TRUE’) and refutation (‘FALSE’) frames by vote preference. Exposure to the confirmation frame has a larger effect among supporters of the incumbent President Alberto Fernandez (*Frente de Todos, FdT*). Supporters of the opposition (*Cambiemos*) and the independent voters (*Blanco*) also display a larger intent to like, share, or comment the confirmation frame, but the effect is more modest.

Table S16 Heterogeneity in effects on reactions in the general specification by partisan attachment and country. Difference of means of reactions by confirmation and refutation frame assignment, partisan attachment, and country.

Variable	Government	Opposition	Blank vote	Opposition v. government
<i>Argentina</i>				
Engage	0.269*** (0.037)	0.160*** (0.027)	0.163*** (0.032)	-0.102** (0.046)
Like	0.238*** (0.034)	0.135*** (0.021)	0.141*** (0.026)	-0.097** (0.039)
Share	0.057** (0.027)	0.042** (0.018)	0.032 (0.021)	-0.010 (0.032)
Comment	0.024 (0.017)	0.004 (0.014)	0.001 (0.013)	-0.023 (0.022)
<i>Brazil</i>				
Engage	0.107*** (0.038)	0.170*** (0.037)	0.099* (0.056)	0.061 (0.053)
Like	0.095*** (0.032)	0.179*** (0.037)	0.098* (0.050)	0.079 (0.049)
Share	0.011 (0.024)	0.000 (0.032)	-0.024 (0.037)	-0.014 (0.039)
Comment	0.023	-0.031	0.002	-0.048

	(0.026)	(0.022)	(0.032)	(0.034)
<i>Chile</i>				
Engage	0.150*** (0.034)	0.163*** (0.040)	0.147*** (0.049)	0.032 (0.052)
Like	0.169*** (0.030)	0.190*** (0.030)	0.155*** (0.038)	0.034 (0.042)
Share	-0.009 (0.023)	0.028 (0.031)	0.021 (0.040)	0.039 (0.039)
Comment	0.004 (0.013)	0.004 (0.020)	0.005 (0.021)	0.001 (0.024)
<i>Colombia</i>				
Engage	0.151*** (0.035)	0.155*** (0.058)	0.124*** (0.043)	0.010 (0.067)
Like	0.115*** (0.029)	0.197*** (0.045)	0.064* (0.034)	0.078 (0.053)
Share	0.040 (0.026)	0.008 (0.051)	0.038 (0.033)	-0.026 (0.056)
Comment	0.032* (0.018)	-0.014 (0.031)	0.030 (0.021)	-0.038 (0.036)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first three columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the second and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S17 Heterogeneity in effects on emotions generation in the general specification by partisan attachment and country. Difference of means of emotions generation by confirmation and refutation frame assignment, partisan attachment, and country.

Variable	Government	Opposition	Blank vote	Opposition v. government
<i>Argentina</i>				
Optimistic	0.377*** (0.034)	0.266*** (0.021)	0.213*** (0.027)	-0.106*** (0.040)
Joyful	0.118*** (0.023)	0.077*** (0.013)	0.066*** (0.015)	-0.045* (0.027)
Angry	-0.177*** (0.027)	-0.115*** (0.019)	-0.086*** (0.020)	0.059* (0.033)
Sad	-0.052***	-0.026**	-0.033**	0.024

	(0.017)	(0.012)	(0.016)	(0.020)
Stressed	-0.102***	-0.094***	-0.017	0.010
	(0.023)	(0.018)	(0.022)	(0.029)
Fearful	-0.018*	-0.011	-0.017	0.010
	(0.011)	(0.012)	(0.017)	(0.017)
Disgusted	-0.159***	-0.118***	-0.095***	0.044
	(0.025)	(0.022)	(0.026)	(0.033)
Indifferent	0.005	-0.014	-0.061	-0.018
	(0.039)	(0.031)	(0.038)	(0.049)
<hr/>				
<i>Brazil</i>				
Optimistic	0.174***	0.343***	0.252***	0.166***
	(0.032)	(0.035)	(0.051)	(0.047)
Joyful	0.071***	0.161***	0.069*	0.090**
	(0.024)	(0.031)	(0.040)	(0.039)
Angry	-0.043*	-0.127***	-0.085***	-0.078**
	(0.024)	(0.021)	(0.030)	(0.031)
Sad	-0.058***	-0.157***	-0.085***	-0.100***
	(0.022)	(0.022)	(0.031)	(0.031)
Stressed	-0.033	-0.093***	0.003	-0.061**
	(0.020)	(0.019)	(0.022)	(0.028)
Fearful	-0.029	-0.072***	-0.054*	-0.036
	(0.024)	(0.020)	(0.030)	(0.031)
Disgusted	-0.034	-0.185***	-0.107***	-0.149***
	(0.021)	(0.024)	(0.030)	(0.032)
Indifferent	-0.067*	0.005	-0.041	0.070
	(0.040)	(0.031)	(0.057)	(0.050)
<hr/>				
<i>Chile</i>				
Optimistic	0.245***	0.303***	0.176***	0.061
	(0.031)	(0.034)	(0.041)	(0.046)
Joyful	0.100***	0.139***	0.075***	0.032
	(0.021)	(0.025)	(0.025)	(0.032)
Angry	-0.096***	-0.069***	-0.069**	0.032
	(0.022)	(0.025)	(0.032)	(0.034)
Sad	-0.013	-0.026*	-0.046**	-0.013
	(0.012)	(0.014)	(0.021)	(0.018)
Stressed	-0.037	-0.033	0.010	0.004
	(0.023)	(0.026)	(0.032)	(0.035)
Fearful	-0.012	-0.075***	-0.057	-0.066**
	(0.022)	(0.025)	(0.034)	(0.032)
Disgusted	-0.069***	-0.028	-0.034	0.045

	(0.023)	(0.026)	(0.034)	(0.035)
Indifferent	-0.092** (0.038)	-0.129*** (0.042)	-0.085 (0.055)	-0.039 (0.056)
<i>Colombia</i>				
Optimistic	0.245*** (0.031)	0.209*** (0.054)	0.176*** (0.038)	-0.015 (0.062)
Joyful	0.088*** (0.019)	0.085** (0.033)	0.032 (0.020)	0.006 (0.038)
Angry	-0.063*** (0.018)	-0.042 (0.030)	-0.036 (0.025)	0.022 (0.036)
Sad	-0.061*** (0.015)	-0.057** (0.024)	-0.019 (0.024)	0.007 (0.029)
Stressed	-0.048** (0.021)	0.016 (0.032)	-0.018 (0.031)	0.059 (0.039)
Fearful	-0.012 (0.025)	-0.084* (0.044)	-0.068* (0.036)	-0.061 (0.050)
Disgusted	-0.018 (0.014)	0.008 (0.025)	0.032 (0.023)	0.035 (0.031)
Indifferent	-0.141*** (0.036)	-0.064 (0.058)	-0.135*** (0.046)	0.052 (0.067)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first three columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the second and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S18 Heterogeneity in levels of reactions in the general specification by partisan attachment and country. Levels of reactions by confirmation and refutation frame assignment, partisan attachment, and country.

Variable	Government		Opposition	
	Refutation	Confirmation	Refutation	Confirmation
<i>Argentina</i>				
Engage	0.223 (0.024)	0.476 (0.029)	0.186 (0.017)	0.346 (0.021)
Like	0.108 (0.018)	0.345 (0.028)	0.073 (0.012)	0.210 (0.018)
Share	0.102	0.145	0.069	0.113

	(0.017)	(0.021)	(0.011)	(0.014)
Comment	0.045 (0.012)	0.066 (0.015)	0.051 (0.010)	0.052 (0.010)
<hr/>				
<i>Brazil</i>				
Engage	0.306 (0.026)	0.406 (0.028)	0.533 (0.027)	0.693 (0.025)
Like	0.164 (0.021)	0.258 (0.025)	0.398 (0.026)	0.561 (0.026)
Share	0.085 (0.016)	0.101 (0.017)	0.219 (0.022)	0.214 (0.022)
Comment	0.110 (0.018)	0.126 (0.019)	0.107 (0.017)	0.087 (0.015)
<hr/>				
<i>Chile</i>				
Engage	0.218 (0.022)	0.368 (0.026)	0.234 (0.025)	0.407 (0.030)
Like	0.117 (0.017)	0.284 (0.024)	0.059 (0.014)	0.255 (0.027)
Share	0.112 (0.017)	0.107 (0.016)	0.134 (0.020)	0.167 (0.023)
Comment	0.032 (0.009)	0.039 (0.010)	0.052 (0.013)	0.057 (0.014)
<hr/>				
<i>Colombia</i>				
Engage	0.284 (0.022)	0.427 (0.026)	0.346 (0.037)	0.507 (0.043)
Like	0.141 (0.017)	0.254 (0.023)	0.086 (0.022)	0.287 (0.039)
Share	0.131 (0.017)	0.172 (0.020)	0.228 (0.033)	0.235 (0.037)
Comment	0.044 (0.010)	0.076 (0.014)	0.080 (0.021)	0.081 (0.023)

Note: Standard errors in parentheses. Each column presents the average value of the reaction indicated in the first column for the restricted sample and the refutation or confirmation frame indicated in the header.

Table S19 Heterogeneity in levels of emotions generation in the general specification by partisan attachment and country. Levels of emotions generation by confirmation and refutation frame assignment, partisan attachment, and country.

Variable	Government		Opposition	
	Refutation	Confirmation	Refutation	Confirmation

<i>Argentina</i>				
Optimistic	0.099 (0.017)	0.469 (0.029)	0.038 (0.008)	0.303 (0.020)
Joyful	0.035 (0.010)	0.155 (0.021)	0.010 (0.004)	0.087 (0.012)
Angry	0.220 (0.023)	0.048 (0.013)	0.158 (0.016)	0.048 (0.009)
Sad	0.073 (0.015)	0.021 (0.008)	0.045 (0.009)	0.022 (0.006)
Stressed	0.137 (0.019)	0.038 (0.011)	0.140 (0.015)	0.048 (0.009)
Fearful	0.035 (0.010)	0.010 (0.006)	0.043 (0.009)	0.032 (0.008)
Disgusted	0.191 (0.022)	0.034 (0.011)	0.206 (0.018)	0.087 (0.012)
Indifferent	0.331 (0.027)	0.328 (0.028)	0.462 (0.022)	0.439 (0.021)

<i>Brazil</i>				
Optimistic	0.117 (0.018)	0.302 (0.026)	0.242 (0.023)	0.583 (0.026)
Joyful	0.063 (0.014)	0.138 (0.019)	0.135 (0.018)	0.290 (0.024)
Angry	0.123 (0.018)	0.072 (0.015)	0.138 (0.019)	0.020 (0.007)
Sad	0.107 (0.017)	0.047 (0.012)	0.170 (0.020)	0.017 (0.007)
Stressed	0.088 (0.016)	0.050 (0.012)	0.118 (0.017)	0.023 (0.008)
Fearful	0.123 (0.018)	0.082 (0.015)	0.104 (0.016)	0.039 (0.010)
Disgusted	0.095 (0.016)	0.057 (0.013)	0.205 (0.022)	0.025 (0.008)
Indifferent	0.451 (0.028)	0.399 (0.028)	0.219 (0.022)	0.223 (0.022)

<i>Chile</i>				
Optimistic	0.117 (0.017)	0.357 (0.025)	0.079 (0.016)	0.380 (0.030)
Joyful	0.034 (0.010)	0.138 (0.018)	0.028 (0.010)	0.160 (0.023)

Angry	0.155 (0.019)	0.053 (0.012)	0.145 (0.021)	0.080 (0.017)
Sad	0.034 (0.010)	0.017 (0.007)	0.045 (0.012)	0.019 (0.008)
Stressed	0.117 (0.017)	0.079 (0.014)	0.124 (0.019)	0.087 (0.017)
Fearful	0.086 (0.015)	0.081 (0.015)	0.138 (0.020)	0.061 (0.015)
Disgusted	0.149 (0.019)	0.076 (0.014)	0.117 (0.019)	0.103 (0.019)
Indifferent	0.473 (0.027)	0.385 (0.026)	0.517 (0.029)	0.388 (0.030)
<hr/>				
<i>Colombia</i>				
Optimistic	0.124 (0.016)	0.362 (0.026)	0.198 (0.031)	0.426 (0.043)
Joyful	0.029 (0.008)	0.113 (0.017)	0.037 (0.015)	0.132 (0.029)
Angry	0.100 (0.015)	0.040 (0.010)	0.099 (0.024)	0.059 (0.020)
Sad	0.075 (0.013)	0.017 (0.007)	0.080 (0.021)	0.029 (0.015)
Stressed	0.109 (0.015)	0.065 (0.013)	0.080 (0.021)	0.096 (0.025)
Fearful	0.146 (0.017)	0.138 (0.018)	0.210 (0.032)	0.140 (0.030)
Disgusted	0.053 (0.011)	0.034 (0.010)	0.049 (0.017)	0.066 (0.021)
Indifferent	0.551 (0.025)	0.407 (0.026)	0.420 (0.039)	0.331 (0.040)

Note: Standard errors in parentheses. Each column presents the average value of the reaction indicated in the first column for the restricted sample and the refutation or confirmation frame indicated in the header.

Table S20 Heterogeneity in effects on reactions in the general specification by vaccination status and country. Difference of means of reactions by confirmation and refutation frame assignment, vaccination status, and country.

Variable	Non-vaccinated [†]	Vaccinated once [†]	Vaccinated twice (or +)
<i>Argentina</i>			
Engage	-0.069	-0.025	0.210***

	(0.073)	(0.094)	(0.019)
Like	-0.034 (0.047)	0.034 (0.076)	0.179*** (0.016)
Share	-0.036 (0.037)	-0.032 (0.067)	0.051*** (0.013)
Comment	-0.003 (0.048)	-0.000 (0.000)	0.007 (0.009)
<hr/>			
<i>Brazil</i>			
Engage	-0.054 (0.115)	0.044 (0.151)	0.151*** (0.025)
Like	-0.024 (0.106)	0.071 (0.130)	0.141*** (0.023)
Share	-0.130 (0.080)	0.046 (0.119)	0.004 (0.019)
Comment	-0.001 (0.045)	-0.210* (0.106)	0.007 (0.016)
<hr/>			
<i>Chile</i>			
Engage	0.010 (0.125)	0.033 (0.178)	0.154*** (0.024)
Like	-0.015 (0.023)	-0.015 (0.198)	0.180*** (0.019)
Share	-0.031 (0.076)	0.088 (0.104)	0.006 (0.017)
Comment	0.055 (0.095)	0.138 (0.270)	0.001 (0.010)
<hr/>			
<i>Colombia</i>			
Engage	-0.038 (0.100)	0.103 (0.077)	0.153*** (0.027)
Like	0.053 (0.035)	0.088 (0.058)	0.122*** (0.022)
Share	-0.064 (0.067)	0.033 (0.053)	0.039* (0.022)
Comment	-0.053 (0.078)	0.044 (0.056)	0.026** (0.012)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

[†] This category comprises around or less than 10% of the observations.

Table S21 Heterogeneity in effects on emotions in the general specification by vaccination status and country. Difference of means of reactions by confirmation and refutation frame assignment, vaccination status, and country.

Variable	Non-vaccinated [†]	Vaccinated once [†]	Vaccinated twice (or +)
<i>Argentina</i>			
Optimistic	0.007 (0.032)	0.089 (0.076)	0.300*** (0.016)
Joyful	0.047 (0.029)	-0.000 (0.000)	0.092*** (0.010)
Angry	-0.086 (0.063)	0.003 (0.077)	-0.131*** (0.013)
Sad	0.032 (0.025)	0.000	-0.042*** (0.009)
Stressed	-0.017 (0.059)	0.042 (0.091)	-0.077*** (0.012)
Fearful	0.025 (0.044)	-0.010 (0.056)	-0.018** (0.008)
Disgusted	-0.205** (0.084)	0.004 (0.092)	-0.123*** (0.014)
Indifferent	0.136 (0.082)	-0.041 (0.116)	-0.034 (0.021)
<i>Brazil</i>			
Optimistic	0.181** (0.075)	0.120 (0.087)	0.270*** (0.023)
Joyful	-0.006 (0.058)	0.047 (0.056)	0.116*** (0.019)
Angry	-0.142* (0.085)	0.026 (0.115)	-0.091*** (0.014)
Sad	-0.024 (0.049)	-0.096 (0.111)	-0.109*** (0.014)
Stressed	-0.051 (0.076)	0.057 (0.121)	-0.056*** (0.012)
Fearful	-0.229*** (0.077)	-0.035 (0.142)	-0.045*** (0.014)
Disgusted	-0.012 (0.095)	-0.030 (0.120)	-0.119*** (0.015)
Indifferent	0.131 (0.118)	-0.113 (0.130)	-0.033 (0.024)
<i>Chile</i>			

Optimistic	0.046 (0.058)	0.177 (0.217)	0.256*** (0.021)
Joyful	0.053 (0.053)	0.167 (0.226)	0.108*** (0.014)
Angry	-0.013 (0.145)	0.222 (0.127)	-0.087*** (0.015)
Sad	0.071 (0.079)	-0.192 (0.213)	-0.026*** (0.009)
Stressed	0.053 (0.100)	-0.202 (0.219)	-0.029* (0.015)
Fearful	-0.099 (0.071)	-0.191 (0.236)	-0.040*** (0.015)
Disgusted	0.224 (0.136)	0.403* (0.183)	-0.059*** (0.015)
Indifferent	-0.046 (0.158)	-0.419 (0.258)	-0.104*** (0.025)
<hr/>			
<i>Colombia</i>			
Optimistic	0.045 (0.066)	0.144** (0.065)	0.240*** (0.024)
Joyful	-0.017 (0.061)	0.027 (0.035)	0.078*** (0.015)
Angry	0.040 (0.068)	0.027 (0.046)	-0.062*** (0.014)
Sad	0.033 (0.065)	-0.020 (0.039)	-0.057*** (0.012)
Stressed	0.050 (0.067)	-0.032 (0.056)	-0.032** (0.016)
Fearful	0.152* (0.083)	-0.015 (0.061)	-0.053*** (0.020)
Disgusted	0.083 (0.076)	0.080* (0.041)	-0.014 (0.011)
Indifferent	-0.226* (0.123)	-0.140* (0.080)	-0.120*** (0.028)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

[†] This category comprises around or less than 10% of the observations.

5.2 Education and attention

In addition to the evidence on heterogeneity by partisan attachment and vaccination status, Tables S22 through S27 add information on the (lack of) heterogeneity in effects by education, time spent reading the question, and attention level. As shown in these tables, there are no significant patterns across countries that support the cognitive hypothesis. As the effect of the confirmation frame is stable across educational attainment on reactions propensity (Table S22) and emotions generation (Table S23), there is no strong evidence that more educated people are less susceptible to the framing¹. Moreover, we compare how the confirmation framing effect varies according to how long do respondents take to answer and how they perform in a simple attention check and find that it is stable. The attention check is defined as successfully solving the operation $(2/2) + 2$ in an independent question. Tables S24 to S27 show that those respondents who spend more time reading the question of interest and those who successfully pass the attention check are either equally or more susceptible to the confirmation frame than those answering more quickly and the ones who fail to solve the mathematical operation. Together, all these results contradict the cognitive burden hypothesis discussed in the main article.

Table S22 Heterogeneity in effects on reactions in the general specification by educational attainment and country. Difference of means of reactions by confirmation and refutation frame assignment, educational attainment, and country.

Variable	Incomplete secondary (or -) [†]	Completed secondary	Incomplete college	Completed college	Incomplete graduate (or +)
<i>Argentina</i>					
Engage	0.309*** (0.074)	0.219*** (0.042)	0.167*** (0.032)	0.166*** (0.031)	0.180** (0.075)
Like	0.151*** (0.055)	0.175*** (0.034)	0.146*** (0.026)	0.170*** (0.026)	0.153** (0.061)
Share	0.142** (0.062)	0.050* (0.028)	0.040* (0.022)	0.021 (0.020)	0.022 (0.052)
Comment	0.046 (0.035)	0.025 (0.021)	-0.002 (0.015)	0.000 (0.014)	0.022 (0.034)
<i>Brazil</i>					
Engage	0.011 (0.083)	0.158*** (0.045)	0.137** (0.064)	0.089* (0.047)	0.196*** (0.052)

¹The Argentina sample is the only one showing a slightly higher effect on engagement for the lowest education category, but this is not significantly different than the effect in any other group.

Like	0.065 (0.071)	0.151*** (0.041)	0.155** (0.060)	0.077* (0.044)	0.176*** (0.051)
Share	-0.019 (0.056)	-0.013 (0.033)	0.026 (0.048)	-0.000 (0.035)	0.004 (0.038)
Comment	-0.039 (0.042)	-0.031 (0.032)	-0.008 (0.040)	0.043 (0.028)	0.002 (0.030)
<hr/>					
<i>Chile</i>					
Engage	0.243 (0.309)	0.157*** (0.052)	0.096* (0.051)	0.179*** (0.034)	0.163** (0.071)
Like	-0.062 (0.223)	0.146*** (0.040)	0.204*** (0.037)	0.178*** (0.028)	0.200*** (0.065)
Share	0.241 (0.356)	-0.002 (0.040)	-0.029 (0.040)	0.040 (0.024)	-0.046 (0.047)
Comment	0.021 (0.085)	0.007 (0.027)	-0.035 (0.026)	0.020 (0.012)	-0.000 (0.029)
<hr/>					
<i>Colombia</i>					
Engage	0.327** (0.147)	0.110** (0.053)	0.154*** (0.050)	0.161*** (0.039)	0.117 (0.078)
Like	0.147 (0.110)	0.159*** (0.042)	0.100** (0.043)	0.098*** (0.031)	0.187*** (0.067)
Share	0.137 (0.157)	-0.032 (0.041)	0.055 (0.036)	0.065** (0.031)	-0.017 (0.071)
Comment	0.086 (0.078)	0.022 (0.030)	0.028 (0.027)	0.021 (0.018)	0.019 (0.028)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

† This category comprises less than 5% of the observations.

Table S23 Heterogeneity in effects on emotions generation in the general specification by educational attainment and country. Difference of means of emotions generation by confirmation and refutation frame assignment, educational attainment, and country.

Variable	Incomplete secondary (or -) [†]	Completed secondary	Incomplete college	Completed college	Incomplete graduate (or +)
<i>Argentina</i>					
Optimistic	0.339*** (0.062)	0.231*** (0.034)	0.321*** (0.027)	0.280*** (0.026)	0.159** (0.061)

Joyful	0.062** (0.028)	0.089*** (0.020)	0.099*** (0.016)	0.053*** (0.016)	0.191*** (0.054)
Angry	-0.026 (0.040)	-0.110*** (0.027)	-0.141*** (0.023)	-0.118*** (0.021)	-0.116*** (0.044)
Sad	-0.076** (0.034)	-0.033 (0.021)	-0.031** (0.013)	-0.032** (0.014)	-0.044* (0.026)
Stressed	-0.036 (0.061)	-0.059* (0.031)	-0.110*** (0.022)	-0.066*** (0.017)	-0.004 (0.035)
Fearful	-0.047 (0.035)	-0.006 (0.022)	-0.003 (0.012)	-0.032*** (0.012)	0.008 (0.009)
Disgusted	-0.131** (0.061)	-0.068** (0.031)	-0.126*** (0.023)	-0.135*** (0.024)	-0.203*** (0.062)
Indifferent	-0.095 (0.086)	-0.025 (0.046)	-0.052 (0.036)	0.000 (0.034)	0.010 (0.084)

Brazil

Optimistic	0.139* (0.072)	0.182*** (0.040)	0.269*** (0.060)	0.289*** (0.041)	0.384*** (0.047)
Joyful	0.092* (0.053)	0.060* (0.032)	0.190*** (0.054)	0.111*** (0.034)	0.123*** (0.042)
Angry	-0.049* (0.028)	-0.081*** (0.025)	-0.096** (0.039)	-0.066** (0.026)	-0.123*** (0.031)
Sad	-0.116** (0.055)	-0.092*** (0.025)	-0.133*** (0.040)	-0.074*** (0.025)	-0.118*** (0.030)
Stressed	-0.005 (0.024)	-0.028 (0.020)	-0.100*** (0.033)	-0.053** (0.026)	-0.060** (0.030)
Fearful	-0.099* (0.058)	-0.012 (0.024)	-0.056 (0.038)	-0.047* (0.024)	-0.093*** (0.032)
Disgusted	-0.064 (0.039)	-0.088*** (0.026)	-0.178*** (0.039)	-0.076*** (0.026)	-0.171*** (0.035)
Indifferent	0.038 (0.086)	-0.029 (0.044)	0.015 (0.062)	-0.066 (0.044)	-0.061 (0.046)

Chile

Optimistic	0.497** (0.178)	0.165*** (0.045)	0.161*** (0.042)	0.317*** (0.030)	0.313*** (0.065)
Joyful	0.100 (0.104)	0.069*** (0.026)	0.096*** (0.028)	0.123*** (0.021)	0.183*** (0.052)
Angry	-0.177 (0.264)	-0.080** (0.034)	-0.086** (0.035)	-0.080*** (0.021)	-0.086** (0.041)
Sad	0.000	-0.026 (0.020)	-0.002 (0.022)	-0.041*** (0.013)	0.000 (0.021)

Stressed	-0.329 (0.238)	-0.006 (0.035)	-0.028 (0.037)	-0.018 (0.022)	-0.046 (0.034)
Fearful	0.140 (0.184)	-0.087** (0.039)	-0.040 (0.033)	-0.042** (0.020)	-0.019 (0.038)
Disgusted	0.002 (0.154)	0.021 (0.031)	-0.032 (0.038)	-0.084*** (0.024)	-0.092** (0.038)
Indifferent	-0.047 (0.391)	-0.023 (0.056)	-0.077 (0.054)	-0.141*** (0.036)	-0.154** (0.074)
<hr/>					
<i>Colombia</i>					
Optimistic	0.265** (0.121)	0.225*** (0.048)	0.284*** (0.044)	0.201*** (0.035)	0.168** (0.077)
Joyful	0.088 (0.073)	0.074** (0.030)	0.068** (0.028)	0.079*** (0.020)	0.024 (0.043)
Angry	0.076 (0.107)	-0.043 (0.027)	-0.033 (0.023)	-0.069*** (0.022)	-0.060 (0.046)
Sad	0.046 (0.080)	-0.039 (0.030)	-0.075*** (0.024)	-0.028* (0.016)	-0.091** (0.036)
Stressed	0.191 (0.120)	0.006 (0.037)	-0.044 (0.029)	-0.047** (0.022)	-0.085 (0.057)
Fearful	0.130 (0.113)	-0.047 (0.046)	-0.100*** (0.036)	0.004 (0.029)	-0.129** (0.053)
Disgusted	0.060 (0.075)	0.035 (0.026)	-0.018 (0.017)	-0.011 (0.017)	-0.047 (0.030)
Indifferent	-0.505*** (0.128)	-0.168*** (0.052)	-0.122** (0.053)	-0.122*** (0.040)	0.062 (0.088)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

† This category comprises less than 5% of the observations.

Table S24 Heterogeneity in effects on reactions in the general specification by quintile of time spent reading the question and country. Difference of means of reactions by confirmation and refutation frame assignment, quintile of time spent reading, and country.

Variable	First quintile	Third quintile	Fifth quintile	Fifth v. first quintile
<i>Argentina</i>				
Engage	0.150*** (0.040)	0.139*** (0.042)	0.247*** (0.043)	0.102* (0.057)

Like	0.141*** (0.034)	0.151*** (0.034)	0.217*** (0.034)	0.089* (0.047)
Share	0.024 (0.025)	0.018 (0.029)	0.080** (0.031)	0.053 (0.039)
Comment	-0.011 (0.016)	0.013 (0.016)	-0.003 (0.025)	0.002 (0.029)
<hr/>				
<i>Brazil</i>				
Engage	0.089 (0.059)	0.076 (0.053)	0.150*** (0.051)	0.072 (0.077)
Like	0.060 (0.054)	0.099* (0.051)	0.091* (0.047)	0.033 (0.070)
Share	0.014 (0.045)	-0.066 (0.040)	0.009 (0.039)	0.005 (0.058)
Comment	-0.021 (0.028)	-0.016 (0.027)	0.043 (0.040)	0.062 (0.049)
<hr/>				
<i>Chile</i>				
Engage	0.235*** (0.055)	0.142*** (0.052)	0.175*** (0.052)	-0.103 (0.074)
Like	0.213*** (0.048)	0.191*** (0.044)	0.120*** (0.040)	-0.117* (0.060)
Share	0.015 (0.037)	0.039 (0.041)	0.030 (0.042)	-0.003 (0.054)
Comment	0.006 (0.020)	0.017 (0.019)	0.066** (0.027)	0.051 (0.033)
<hr/>				
<i>Colombia</i>				
Engage	0.201*** (0.067)	0.062 (0.057)	0.119** (0.052)	-0.081 (0.081)
Like	0.189*** (0.057)	0.083* (0.046)	0.080* (0.041)	-0.086 (0.067)
Share	0.047 (0.051)	-0.021 (0.043)	0.076* (0.043)	0.027 (0.063)
Comment	-0.019 (0.029)	0.010 (0.024)	0.025 (0.031)	0.020 (0.042)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first three columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the third and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S25 Heterogeneity in effects on emotions generation in the general specification by quintile of time spent reading the question and country. Difference of means of emotions generation by confirmation and refutation frame assignment, quintile of time spent reading, and country.

Variable	First quintile	Third quintile	Fifth quintile	Fifth v. first quintile
<i>Argentina</i>				
Optimistic	0.206*** (0.032)	0.279*** (0.037)	0.371*** (0.036)	0.157*** (0.048)
Joyful	0.084*** (0.021)	0.075*** (0.021)	0.062*** (0.020)	-0.017 (0.029)
Angry	-0.111*** (0.028)	-0.099*** (0.027)	-0.121*** (0.031)	-0.010 (0.040)
Sad	-0.032* (0.019)	-0.028* (0.015)	-0.030 (0.020)	-0.003 (0.027)
Stressed	-0.059** (0.026)	-0.125*** (0.032)	-0.090*** (0.027)	-0.030 (0.036)
Fearful	-0.016 (0.016)	-0.020 (0.020)	-0.017 (0.019)	0.002 (0.024)
Disgusted	-0.120*** (0.033)	-0.073** (0.029)	-0.155*** (0.032)	-0.032 (0.045)
Indifferent	0.021 (0.047)	-0.034 (0.047)	-0.061 (0.044)	-0.085 (0.064)
<i>Brazil</i>				
Optimistic	0.182*** (0.049)	0.289*** (0.048)	0.226*** (0.048)	0.056 (0.067)
Joyful	0.067* (0.038)	0.079* (0.043)	0.132*** (0.037)	0.070 (0.053)
Angry	-0.080*** (0.030)	-0.063** (0.031)	-0.077*** (0.029)	0.015 (0.042)
Sad	-0.095*** (0.033)	-0.065** (0.025)	-0.111*** (0.033)	-0.023 (0.047)
Stressed	-0.083** (0.036)	-0.027 (0.026)	-0.015 (0.025)	0.062 (0.042)
Fearful	-0.053* (0.031)	-0.019 (0.026)	-0.054 (0.036)	0.012 (0.047)
Disgusted	-0.141*** (0.033)	-0.131*** (0.031)	-0.137*** (0.033)	0.005 (0.046)
Indifferent	0.039 (0.056)	-0.036 (0.050)	-0.004 (0.048)	-0.070 (0.073)

<i>Chile</i>				
Optimistic	0.257*** (0.047)	0.271*** (0.048)	0.244*** (0.047)	-0.031 (0.064)
Joyful	0.135*** (0.033)	0.126*** (0.031)	0.157*** (0.037)	0.012 (0.048)
Angry	-0.065* (0.033)	-0.053 (0.033)	-0.108*** (0.033)	-0.036 (0.047)
Sad	-0.020 (0.015)	-0.014 (0.019)	-0.024 (0.019)	-0.005 (0.024)
Stressed	-0.086** (0.036)	0.022 (0.030)	0.011 (0.034)	0.080* (0.049)
Fearful	-0.060** (0.030)	-0.014 (0.032)	-0.029 (0.037)	0.013 (0.046)
Disgusted	-0.051 (0.034)	-0.026 (0.034)	-0.027 (0.036)	0.011 (0.049)
Indifferent	-0.113* (0.061)	-0.157*** (0.057)	-0.138*** (0.053)	0.005 (0.079)

<i>Colombia</i>				
Optimistic	0.211*** (0.059)	0.204*** (0.052)	0.174*** (0.047)	-0.019 (0.071)
Joyful	0.049 (0.034)	0.029 (0.025)	0.067** (0.030)	0.008 (0.044)
Angry	-0.048* (0.029)	-0.066** (0.028)	-0.047 (0.030)	-0.006 (0.041)
Sad	0.002 (0.025)	-0.073*** (0.024)	-0.047 (0.030)	-0.043 (0.041)
Stressed	0.048 (0.044)	-0.037 (0.034)	-0.029 (0.034)	-0.066 (0.053)
Fearful	-0.025 (0.045)	-0.069* (0.039)	-0.026 (0.043)	0.006 (0.060)
Disgusted	0.001 (0.019)	-0.033 (0.028)	0.040* (0.022)	0.057* (0.030)
Indifferent	-0.182*** (0.067)	0.009 (0.059)	-0.149*** (0.052)	0.019 (0.081)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first three columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the third and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex,

educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S26 Heterogeneity in effects on reactions in the general specification by fulfillment of attention check and country. Difference of means of reactions by confirmation and refutation frame assignment, fulfillment of attention check, and country.

Variable	Inattentive	Attentive	Difference
<i>Brazil</i>			
Engage	0.112** (0.045)	0.138*** (0.028)	0.020 (0.053)
Like	0.104** (0.041)	0.134*** (0.026)	0.026 (0.049)
Share	-0.001 (0.033)	0.001 (0.021)	0.002 (0.039)
Comment	-0.008 (0.030)	0.004 (0.017)	0.008 (0.035)
<i>Chile</i>			
Engage	0.174*** (0.044)	0.145*** (0.027)	-0.021 (0.051)
Like	0.157*** (0.034)	0.177*** (0.022)	0.031 (0.040)
Share	0.015 (0.032)	0.006 (0.020)	-0.008 (0.037)
Comment	-0.008 (0.022)	0.010 (0.011)	0.014 (0.024)
<i>Colombia</i>			
Engage	0.189*** (0.049)	0.141*** (0.028)	-0.024 (0.056)
Like	0.129*** (0.039)	0.114*** (0.023)	-0.013 (0.045)
Share	0.020 (0.041)	0.045** (0.022)	0.034 (0.045)
Comment	0.098*** (0.028)	0.006 (0.014)	-0.084*** (0.031)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first two columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the reaction indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the second and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

Table S27 Heterogeneity in effects on emotions generation in the general specification by fulfillment of attention check and country. Difference of means of emotions generation by confirmation and refutation frame assignment, fulfillment of attention check, and country.

Variable	Inattentive	Attentive	Difference
<i>Brazil</i>			
Optimistic	0.187*** (0.040)	0.293*** (0.026)	0.114** (0.047)
Joyful	0.061** (0.030)	0.126*** (0.022)	0.068* (0.037)
Angry	-0.097*** (0.023)	-0.079*** (0.017)	0.014 (0.028)
Sad	-0.108*** (0.028)	-0.098*** (0.015)	0.006 (0.031)
Stressed	-0.011 (0.021)	-0.068*** (0.015)	-0.057** (0.025)
Fearful	-0.032 (0.025)	-0.057*** (0.016)	-0.023 (0.030)
Disgusted	-0.106*** (0.025)	-0.117*** (0.018)	-0.014 (0.030)
Indifferent	-0.016 (0.043)	-0.035 (0.027)	-0.025 (0.051)
<i>Chile</i>			
Optimistic	0.265*** (0.037)	0.243*** (0.024)	-0.013 (0.043)
Joyful	0.100*** (0.025)	0.107*** (0.016)	0.008 (0.029)
Angry	-0.089*** (0.030)	-0.081*** (0.017)	0.006 (0.035)
Sad	-0.020 (0.017)	-0.023** (0.010)	-0.002 (0.020)
Stressed	-0.045 (0.030)	-0.017 (0.017)	0.024 (0.035)
Fearful	-0.080*** (0.031)	-0.029* (0.017)	0.054 (0.034)
Disgusted	-0.013 (0.033)	-0.061*** (0.017)	-0.057 (0.037)
Indifferent	-0.097** (0.049)	-0.103*** (0.029)	-0.013 (0.056)
<i>Colombia</i>			

Optimistic	0.282*** (0.044)	0.209*** (0.025)	-0.069 (0.050)
Joyful	0.073*** (0.026)	0.068*** (0.015)	-0.010 (0.029)
Angry	-0.017 (0.022)	-0.059*** (0.015)	-0.037 (0.028)
Sad	-0.077*** (0.025)	-0.038*** (0.013)	0.044 (0.028)
Stressed	-0.007 (0.031)	-0.037** (0.017)	-0.027 (0.036)
Fearful	-0.026 (0.042)	-0.041** (0.020)	-0.003 (0.046)
Disgusted	0.006 (0.020)	-0.003 (0.013)	-0.007 (0.025)
Indifferent	-0.155*** (0.051)	-0.121*** (0.029)	0.038 (0.057)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each cell in the first two columns corresponds to a different regression using only the restricted sample indicated in the header. Coefficients represent the effect of the confirmation frame on the emotion indicated in the first column compared against the refutation frame. The last column present the difference between coefficients in the second and first columns from the interaction of the treatment with the categories indicated. All regressions control for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post.

6 Placebo experiment

In addition to the vaccine framings, a placebo experiment was conducted using a picture of a dog and varying only the first sentence of the statement: “It is TRUE” or “It is FALSE”. The second sentence was common to both framings and read “Study shows that dogs do not really know what you are saying to them” (see Figure S6). Tables S28 to S30 present the results for the placebo framings for Brazil, Chile, and Colombia, respectively. In all cases, albeit attenuated, there is still a positive effect of the confirmation frame on engagement through likes, but all the other dependent variables are found to be framing invariant. This result, together with the heterogeneity by partisan attachment presented in section 5.1, can be interpreted as showing that framing effects depend on how salient the issue is to the reader and how it reinforces or contradicts an ideological prior rather than a non-ideological fact.

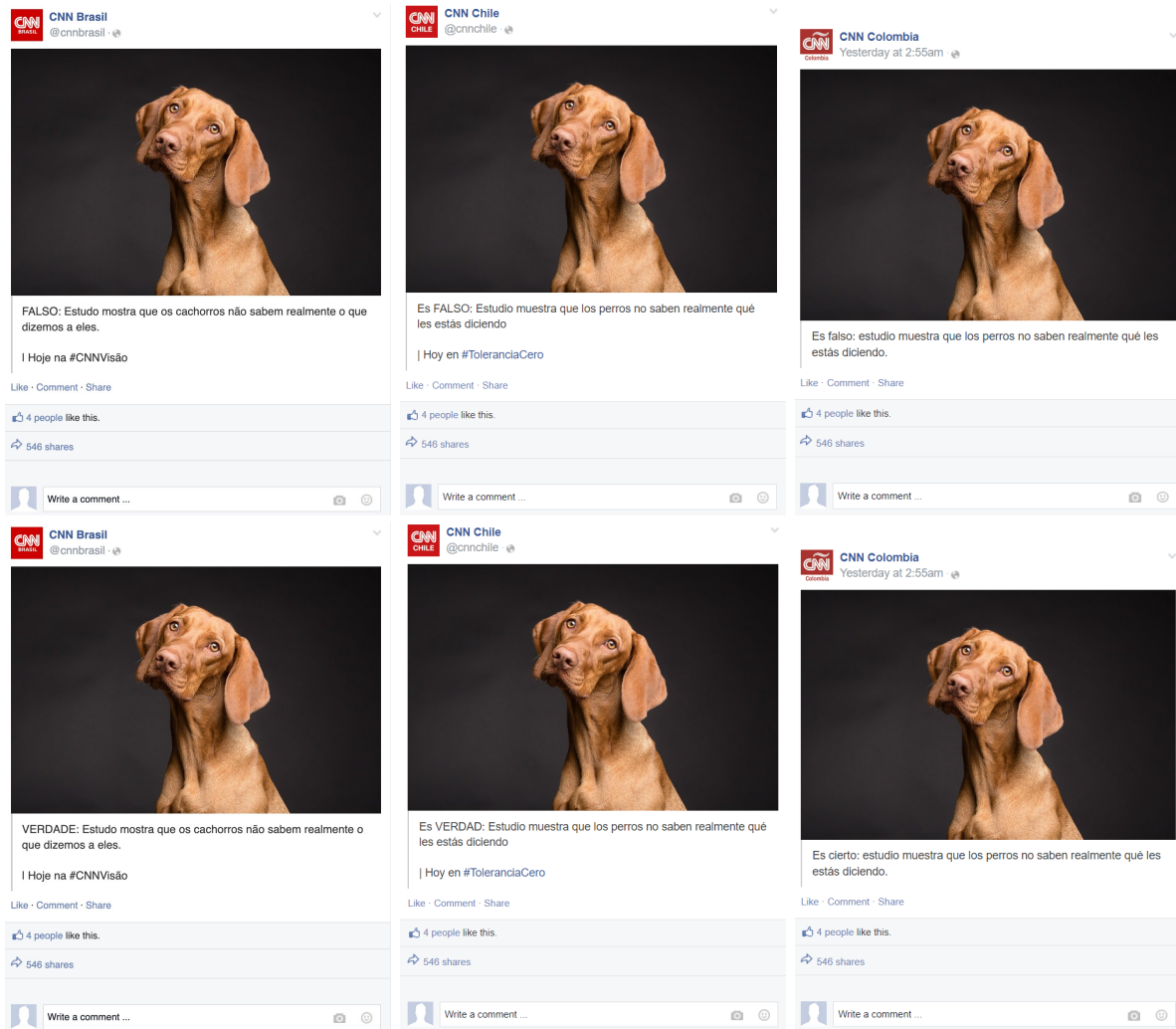


Figure S6 Images of the *confirmation* and *refutation* placebo treatments used in Brazil, Chile, and Colombia. The confirmation and refutation frames are semantically equivalent and intended to be equivalent in their cognitive accessibility and their valence charge, changing only the word “True” for “False”. The texts read “It is [TRUE/FALSE]. Study shows that dogs do not really know what you are saying to them”. All four treatments are factually correct and conform to the design used by our partner organization in Argentina, *Chequeado*.

Table S28 Difference of means between the confirmation and refutation frames in the placebo (dog) experiment in the Brazil survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.354 (0.025)	0.431 (0.025)	0.077** (0.036)	0.066* (0.036)
Like	0.248 (0.023)	0.317 (0.024)	0.069** (0.033)	0.063* (0.033)
Share	0.076 (0.014)	0.086 (0.014)	0.009 (0.020)	0.007 (0.020)
Comment	0.120 (0.017)	0.114 (0.016)	-0.006 (0.023)	-0.012 (0.024)
<i>Emotions</i>				
Optimistic	0.087 (0.015)	0.086 (0.014)	-0.001 (0.021)	-0.007 (0.021)
Joyful	0.082 (0.014)	0.068 (0.013)	-0.014 (0.019)	-0.016 (0.019)
Angry	0.054 (0.012)	0.042 (0.010)	-0.013 (0.016)	-0.012 (0.016)
Sad	0.125 (0.017)	0.096 (0.015)	-0.029 (0.023)	-0.028 (0.023)
Stressed	0.030 (0.009)	0.031 (0.009)	0.001 (0.013)	-0.003 (0.013)
Fearful	0.027 (0.009)	0.010 (0.005)	-0.017* (0.010)	-0.019* (0.010)
Disgusted	0.068 (0.013)	0.081 (0.014)	0.012 (0.019)	0.009 (0.019)
Indifferent	0.629 (0.025)	0.642 (0.024)	0.012 (0.035)	0.022 (0.035)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively in the placebo experiment. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

Table S29 Difference of means between the confirmation and refutation frames in the placebo (dog) experiment in the Chile survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.354 (0.025)	0.434 (0.025)	0.080** (0.035)	0.084** (0.036)
Like	0.192 (0.021)	0.267 (0.022)	0.075** (0.030)	0.071** (0.031)
Share	0.109 (0.016)	0.125 (0.017)	0.016 (0.023)	0.021 (0.023)
Comment	0.148 (0.019)	0.142 (0.017)	-0.005 (0.026)	0.007 (0.026)
<i>Emotions</i>				
Optimistic	0.089 (0.015)	0.127 (0.017)	0.038* (0.022)	0.045** (0.023)
Joyful	0.081 (0.014)	0.090 (0.014)	0.009 (0.020)	0.012 (0.020)
Angry	0.109 (0.016)	0.082 (0.014)	-0.026 (0.021)	-0.025 (0.022)
Sad	0.134 (0.018)	0.150 (0.018)	0.016 (0.025)	0.009 (0.026)
Stressed	0.036 (0.010)	0.022 (0.007)	-0.014 (0.012)	-0.013 (0.013)
Fearful	0.019 (0.007)	0.015 (0.006)	-0.005 (0.009)	-0.004 (0.010)
Disgusted	0.031 (0.009)	0.025 (0.008)	-0.006 (0.012)	-0.005 (0.012)
Indifferent	0.610 (0.026)	0.586 (0.025)	-0.024 (0.036)	-0.021 (0.036)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively in the placebo experiment. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.

Table S30 Difference of means between the confirmation and refutation frames in the placebo (dog) experiment in the Colombia survey.

Variable	Refutation frame	Confirmation frame	Simple difference	Difference with controls
<i>Reactions</i>				
Engage	0.440 (0.025)	0.529 (0.025)	0.089** (0.035)	0.087** (0.035)
Like	0.246 (0.021)	0.267 (0.022)	0.021 (0.030)	0.020 (0.031)
Share	0.123 (0.016)	0.192 (0.019)	0.069*** (0.025)	0.067*** (0.025)
Comment	0.162 (0.018)	0.190 (0.019)	0.028 (0.027)	0.028 (0.027)
<i>Emotions</i>				
Optimistic	0.118 (0.016)	0.156 (0.018)	0.038 (0.024)	0.038 (0.024)
Joyful	0.106 (0.015)	0.123 (0.016)	0.017 (0.022)	0.016 (0.022)
Angry	0.076 (0.013)	0.125 (0.016)	0.049** (0.021)	0.048** (0.021)
Sad	0.147 (0.018)	0.147 (0.017)	-0.001 (0.025)	0.004 (0.025)
Stressed	0.015 (0.006)	0.012 (0.005)	-0.003 (0.008)	-0.000 (0.007)
Fearful	0.017 (0.006)	0.026 (0.008)	0.009 (0.010)	0.015 (0.010)
Disgusted	0.022 (0.007)	0.017 (0.006)	-0.005 (0.010)	-0.008 (0.010)
Indifferent	0.592 (0.024)	0.536 (0.024)	-0.056 (0.035)	-0.057* (0.034)

Note: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The first two columns represent average values for each reaction and emotion in response to the refutation and confirmation frames respectively in the placebo experiment. The third column presents the difference between the confirmation and refutation frame without controls. Differences in the fourth column are estimated controlling for age, sex, educational attainment, employment status, partisan attachment, having had COVID-19, number of doses administered of COVID-19 vaccine, and time spent reading the post. Significance levels are derived from linear regression models.