

# **Doubt the Messenger: The reputation cost of fact-checking**

Natalia Aruguete\*      Ernesto Calvo †      Tiago Ventura ‡

March 24, 2022

## **Abstract**

We implement a survey experiment to measure the reputation cost of publishing pro- and counter-attitudinal fact checks in a polarized political environment. The survey experiment exposes respondents to a news publication that was TRUE or FALSE depending on the day of its publication. Later in the survey we treat respondents to a fact check that adjudicates the initial tweet as TRUE or FALSE. Respondents indicate a higher intent to share pro-attitudinal fact checks compared to counter-attitudinal fact checks. Concurrently, the treated group answered questions about the perceived reputation of a menu of news organizations, politicians, and the fact checker “Chequeado”. A control group answered those same reputation questions before observing the correction by *Chequeado*. The experiment shows that a counter-attitudinal adjudication reduces the perceived reputation of the fact checker compared to the pro-attitudinal adjudication. The counter-attitudinal adjudication also increases the perceived ideological distance between the respondent and the fact checker (contrast effect). Results carry important implications for the long term sustainability of organizations interested in reducing the spread of misinformation.

---

\*Universidad Nacional de Quilmes, UNQ. Email: [nataliaaruguete@gmail.com](mailto:nataliaaruguete@gmail.com). Webpage: <http://unq.academia.edu/nataliaaruguete>

†University of Maryland, Government and Politics, UMD. Address: 3140 Tydings Hall, College Park, MD 20742, USA. Email: [ecalvo@umd.edu](mailto:ecalvo@umd.edu). Webpage: <http://gvptsites.umd.edu/calvo/>

‡University of Maryland, Government and Politics, UMD. Address: 4118 Chiconteague, College Park, MD 20742, USA. Email: [venturat@umd.edu](mailto:venturat@umd.edu). Webpage: <http://tiagoventura.rbind.io/>

# 1 Reputation and Misinformation

## 1.1 Introduction

Political fact-checking has become a central effort against the prevalence of misinformation. 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), fact-checking is a global phenomenon, with scores of initiatives spearheaded by news organizations, independent media, and NGOs (Graves, 2018). The popularity of fact-checking stems, in part, from its efficacy as a remedy against misinformation. A recent meta-analysis found that fact-checking messages are successful at reducing misperceptions, even after a single exposure (Walter et al., 2020, p.350). However, the same meta-analysis found that this positive effect is conditioned by context, audience, and message characteristics.

The current article expands on the literature of moderators of fact-checking, focusing on the reputation costs faced by the fact checker. Reputation is critical to the mission of the fact checker, affecting the long term viability of its interventions, which require readers to accept that adjudications of TRUE or FALSE are valid. 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). Emulating a judicial process, the act of adjudicating content as TRUE or FALSE repairs the damage inflicted upon an injured group and imposes a cost on the offending one.<sup>1</sup>

---

<sup>1</sup>For example, reporting that “It is FALSE that vaccines do not prevent against COVID-19” will damage the group spreading vaccine misinformation and benefit the group that supports vaccination mandates. The use of the term *adjudication*, rather than the more frequent term *correction*, highlights the confrontational nature of the fact checking.

An important condition for a successful adjudication is that the equanimity and unbiasedness of the fact checker is widely recognized. This reputation is integral to the adjudication process. Further, the reputation of the fact checker is put to the test every time they adjudicate information as correct or incorrect, TRUE or FALSE. Indeed, there is a wealth of research showing that counter-attitudinal adjudications of TRUE or FALSE are often met with disbelief by users, who challenge both the intent of the correction and the factual accuracy of the fact check. Accordingly, there is a reputation capital that may be lost when adjudicating content as factually correct or as factually incorrect. This raises the possibility that future interventions will be less effective and that, over time, the perceived quality and ideological integrity of the fact checker will decline, threaten its viability as an effective organization.

This article presents results of a survey experiment measuring the reputation cost of publishing pro- and counter-attitudinal fact checks in a politically charged environment. The survey experiment exposes respondents to one of two real tweets posted by a leading Argentine news organization, *Infobae*, which states that Argentina had the highest number of COVID-19 cases per day on the previous day. Both tweets were published six months apart by *Infobae* with an identical text. An October 2, 2020 publication was adjudicated as FALSE by the leading Argentine fact checker, Chequeado, and posted in their website. By contrast, an identically worded article published in May 20, 2021 was factually correct. Content that may be accurate or inaccurate depending on the day provide researchers with an opportunity to assess the effect of fact checks without requiring deception at the time of implementation.

The publication by *Infobae* aligns well with charges frequently leveled against Argentine President Alberto Fernandez by the opposition, who accused his administration of under performing during the pandemics. Accordingly, we expect opposition voters to more widely like and share

the original tweet compared to respondents favorable to the government. In our survey, conducted a few days before the 2021 election, respondents were randomly assigned to a control group, a TRUE group, and a FALSE group. After respondents were treated to one of the two original Infobae tweets, but prior to exposure to Chequeado's adjudication, the control group answered questions about the perceived reputation and ideological position of a variety of news organizations and politicians, among which we include the fact checker *Chequeado*. The treated groups, on the other hand, were presented with either a TRUE or a FALSE adjudication before answering the reputation and ideology questions. Therefore, the experiment tests for the perceived reputation of the fact checker among those not exposed to an adjudication, among those exposed to an adjudication of TRUE, and among those exposed to an adjudication of FALSE.

Result from our experiment show that respondents perceived *Chequeado* as more reputable when observing the pro-attitudinal adjudication (i.e. the TRUE adjudication for opposition supporters and the FALSE adjudication for government supporters) compared to the counter-attitudinal adjudication (i.e. the FALSE adjudication for opposition supporters and the TRUE adjudication for government supporters). Respondents also perceived Chequeado further distant ideologically (contrast effect) when exposed to counter-attitudinal adjudications compared to pro-attitudinal ones. However, this contrast effect was statistically significant only for supporters of the government.

The results of our analyses carry important implications for understanding the reputation costs of counter-attitudinal adjudications. As noted by Brandtzaeg and Følstad (2017)[.67]“The more political or controversial issues a fact-checking service covers, the more it needs to build a reputation for usefulness and trustworthiness.” While the decision to correct misinformation cannot be subject to an instrumental calculus of its reputation costs and benefits, our findings

do raise the question of how to better mitigate reputation costs over time. This may include making explicit and publicizing the editorial policies used to select fact checks, to describe the rules used to report content as misinformation, and to strategically frame adjudications to both curve misinformation and maintain a healthy reputation to ensure future fact check interventions.

## 1.2 Reputation as Quality and Equanimity

An ever expanding literature shows that ‘partisanship is inextricably linked with social identities’(Mason et al., 2021). These partisan identities “color one’s interpretation of political information”(Bolsen et al., 2014, 236), including our perceptions of the out-group voters, political parties, and news organizations (Ahler and Sood, 2018; Broockman and Skovron, 2018; Ardèvol-Abreu and Gil de Zúñiga, 2017). Increasingly, the literature describes voters as “motivated reasoners who seek out congenial sources of information and defend their attitudes and beliefs when challenged”(Strickland et al., 2011, 935). In a series of landmark studies, Lodge and Taber (2013) demonstrate that “motivated reasoning - the systematic biasing of judgements in favor of automatically activated, affectively congruent beliefs and feelings - is built into the basic architecture and information processing of the brain”(Lodge and Taber, 2013, 24).<sup>2</sup>

The effect of motivated reasoning on the perceived quality and ideological leaning of news organizations has been well documented in the communications’ literature (Ardèvol-Abreu and Gil de Zúñiga, 2017; Lee, 2005, 2012). Since it was first discussed by Vallone et al. (1985), a significant scholarship refers to this negative partisan evaluations as a “hostile media” effect (Vallone et al., 1985; Gunther and Schmitt, 2004; Hansen and Kim, 2011), whereby readers

<sup>2</sup>The extant literature uses a number of different terms to describe misperceptions in the location of parties, including *perceptual bias* (??Aldrich et al., 2018), *projection bias* (Muraoka and Rosas, 2021), *halo effect* (Kahneman, 2011), and *assimilation and contrast* (Adams et al., 2005). We will restrict the terminology of our article to *perceptual bias* for the general effect and to *assimilation and contrast* for the positive and negative perceptual biases respectively.

perceive news organizations to be ideologically distant and biased against them. In their classic study of the Israeli-Palestinian conflict coverage, [Vallone et al. \(1985\)](#) note that: “First, partisans evaluated the fairness of the media’s sample of facts and arguments differently: in light of their own divergent views about the objective merits of each side’s case and their corresponding views about the nature of unbiased coverage. Second, partisans reported different perceptions and recollections about the program content itself; that is, each group reported more negative references to their side than positive ones, and each predicted that the coverage would sway nonpartisans in a hostile direction.”

Assimilation and contrast effects were concurrently studied by political scientists, although research focused on the perceived ideology of political parties and to a lesser extent of news organizations ([Iyengar and Kinder, 2010](#); [Adams et al., 2005](#); [Broockman and Skovron, 2018](#)). While different in scope and size from the traditional news organizations, fact checkers are today integral to the news media environment. Fact checkers are frequently cited as authorities to adjudicate misinformation intent to partisans during elections and often face significant backlash from the affected groups that amplified misinformation content. To our knowledge, however, there is little research that has explored how assimilation and contrast effects moderate the perceived ideological leaning and the perceived reputation of the fact checkers.

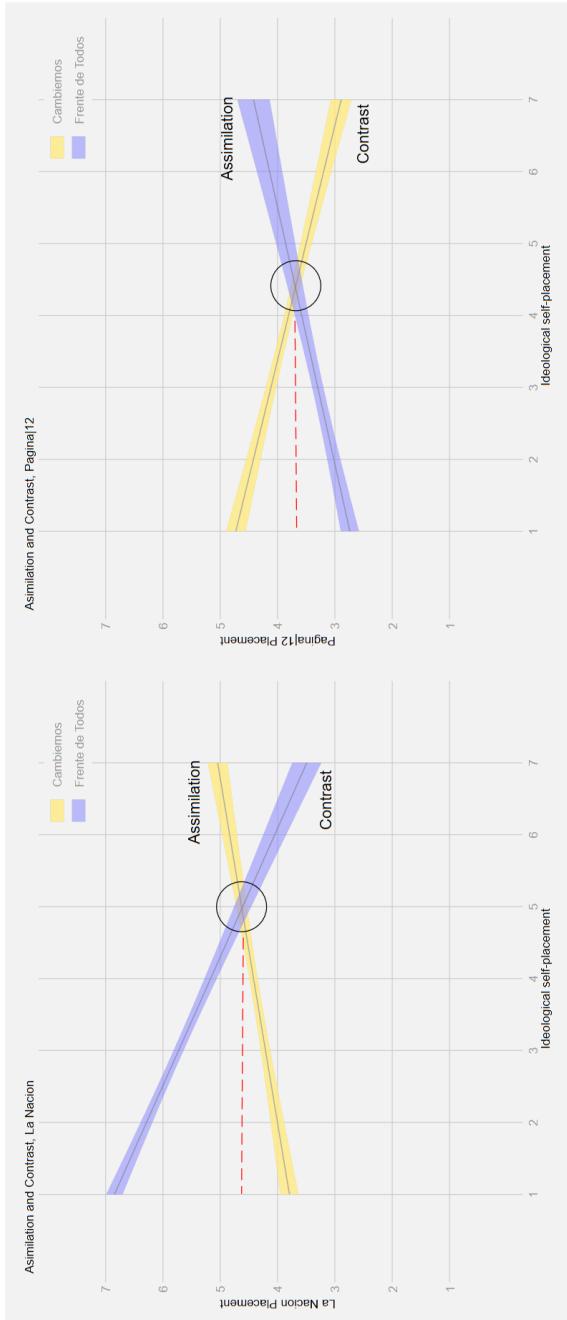
### **1.3 Assimilation and Contrast in Perceived Media Bias**

*Assimilation* and *contrast* describes positive and negative valence charges that bias our judgement when judging the position of individuals, organizations, or objects. Assimilation describes the tendency to perceive that these entities are ideologically closer to us. Contrast, on the other hand, biases the perceived position of a news organization which is observed as further removed.

*Assimilation* and *contrast* effects are pervasive in survey data. [Merrill et al. \(2001\)](#) analyze elections in Norway, France, and the United States and describe consistent and general assimilation and contrast effects for all parties and voters in each of them. They estimate significant perceptual bias in the random placement of parties around a latent “true” mean (stochastic variation) as well as distortions in the scale of the space (anchoring effects). Calvo, Chang, and Hellwig ([2014](#)) study 83 parties in 13 countries with data from the Comparative Study of Electoral Systems (CSES) and find large and statistically significant perceptual biases in 82 of them. Both [Aldrich et al. \(2018\)](#) and [Muraoka and Rosas \(2021\)](#) analyze individual, economic, and party system determinants of perceptual bias in over 300 parties, showing higher *deviations* in older democracies, majoritarian electoral systems with presidential elections. In the United States, one of those older, majoritarian, and presidential regimes, the assimilation and contrast curves are extraordinarily stable over 50 years of the National Election Study data ([Calvo et al., 2021](#)).<sup>3</sup>

---

<sup>3</sup>In the early 1970s and 1980s there are a number of studies that documented the prevalence of assimilation and contrast in US politics. Interestingly, most survey data showed large assimilation and modest contrast effects, as documented in the meta analysis of [Granberg et al. \(1981\)](#). Since the 1990s, on the other hand, assimilation has declined and contrast has increased in importance. In their 1981 article, [Granberg et al. \(1981\)](#) noted: “It is still not known why a candidate is assimilated or contrasted to different degrees on different issues.” After the 1980s, despite how well documented are these effects, research interest declined. We believe that was premature and that modeling perceptual bias explicitly is an important research agenda



*Note:* Estimates from the Argentine National Election Survey (2021). The horizontal axis describes the self-reported location of the respondents. The vertical axis describes the reported location of the conservative news paper *La Nación* (left) and the progressive newspaper *Página 12* (right). Blue lines describe how government supporters of the *Frente de Todos* perceive each newspaper and yellow lines describe how supporters of the opposition *Cambiemos* perceive each newspaper. Red lines describe the location at which each news organization would be perceived as identical by supporters of both parties. This “unbiased” user location is modestly conservative for *La Nación* and modestly liberal for *Página 12*. The contrast line for *La Nación* shows that as users become more liberal their perceive the newspaper as more conservative. On the other hand, supporters of *Cambiemos* always perceive *La Nación* as ideologically closer, therefore the positive slope. In the case of *Página 12*, conservative voters perceive the newspaper increasingly progressive while government supporters see a positive slope.

Let us consider as an example two major news organizations, one widely perceived as progressive and aligned with the current administration of peronist President Alberto Fernandez, Pagina|12, and another one widely perceived as conservative and aligned with the opposition led by former President Mauricio Macri, La Nación. Using data from the Argentine National Election Survey (2021), Figure 1 describes how the progressive or conservative ideology of the survey respondents alters the perceived location of the news organization.<sup>4</sup> The horizontal axis describes the self-reported location of the respondents in a scale that ranges from 1 (very progressive) to 7 (very conservative). The vertical axis describes the same respondents' perceived location of the conservative newspapers La Nación (left plot) and the progressive newspaper Pagina|12 (right plot). Blue lines (linear specification) describe how government supporters of the *Frente de Todos* perceive each newspapers and yellow lines describe how supporters of the opposition coalition *Cambios* perceive each newspaper. Red dotted lines describe the location at which each news organization would be perceived as identical by supporters of both parties.

This *locally unbiased ideological location*, as described in the next section, is modestly conservative for La Nación and modestly liberal for Pagina|12. However, there are few respondents at this “unbiased” locations. Most supporters see Pagina|12 and La Nación further to the left or right. The contrast line for La Nación (blue) shows that as the Frente de Todos’ users become more progressive they perceive the the newspaper as more conservative. The assimilated supporters of Cambios, on the other hand, perceive La Nación as ideologically closer as they move to the right.

---

<sup>4</sup>Each lines describes a separate OLS model with a dependent variable that reads: “Imagine a scale from ‘VERY PROGRESSIVE’ to ‘VERY CONSERVATIVE’. Where would you place the newspaper La Nación?”. Responses take values ranging from 1 (very progressive) to 7 (very conservative). The independent variable asks the same question about the respondent (self-reported conservatism). For example, the contrast line for La Nación is estimated as  $y_{fdt} = 7.406 - .559x_{fdt}$ , showing that each unit of increase in conservatism for an Frente de Todos voter results in a decline in conservatism for La Nación of -.559. In other words, the more progressive the supporter of the Frente de Todos, the more conservative they perceive La Nacion to be. On the other hand, Cambios supporters are assimilated, which is reflected in a positive slope,  $y_{fdt} = 3.58 + .209x_{fdt}$ .

## 1.4 Position and Valence Shocks

Assimilation and contrast have a statistical interpretation as the result of a mixture of positional information (left-right) and valence shocks (positive-negative). Consider a newspaper that is widely read only by supporters of a progressive incumbent president, such as Pagina|12. The newspaper produces news articles that are positioned to the left of the political spectrum and, concurrently, these articles are read by partisans that more frequently interact with one another (i.e. in-group peers). The editorial position of the newspaper is expected to align with the in-group readers, adding a positive shock that alters its perceived position. Positive valence shocks result in a weighted ideological position that lies between the latent editorial ideology of the newspaper and the assimilated ideological position of the reader. A negative valence shock, on the other hand, pushes the perceived position of the newspaper away from the voter and beyond the latent editorial position of Pagina|12 (contrast).

A formal description of this aggregate bias was recently derived by Calvo, Chang, and Dodyk (2022). Consider a reader with an ideology position  $x$  in a latent progressive to conservative dimension. The reader perceives the ideology of an in-group favored newspaper as:

$$A(x) := \alpha_A + \beta_A x,$$

with  $\beta_A > 0$  (*assimilation*). If the individual is not aligned with the majority of the newspaper's readership, on the other hand, her perception of the newspaper's ideology is:

$$C(x) := \alpha_C + \beta_C x,$$

with  $\beta_C < 0$  (*contrast*).

A news organization will be globally unbiased if its coverage of a news event also satisfies the condition  $\beta_A = \beta_C = 0$  for the reader  $x$ . This unbiasedness does not require a news paper to be centrist. Rather, unbiasedness indicates that there is no in-group or out-group effect in the reported locations, which results in all readers observing the mean reported location (with some random stochastic variation around its mean). In the presence of assimilation and contrast, however, the news organization will be perceived as favorable biased by in-group supporters and unfavorably biased by the out-group supporters.

If assimilation and contrast dominate the data, we may still derive a locally *unbiased* reader whose perception of the ideological leaning of the news organization is independent of her own alignment with the majority of its in-group readers. This unbiased voter is still a motivated reasoner, just as the other voters in the sample, but the motivated search for evidence converges to an identical location when switching from the in-group to the out-group. In Figure 1, the location of this locally unbiased voter is represented with a circle exactly where the linear model isolines estimated in the assimilation and contrast equations intersect.

Solving for  $x$ , the position of the locally unbiased reader we obtain:

$$x^{\text{unbiased}} := -\frac{\alpha_A - \alpha_C}{\beta_A - \beta_C}.$$

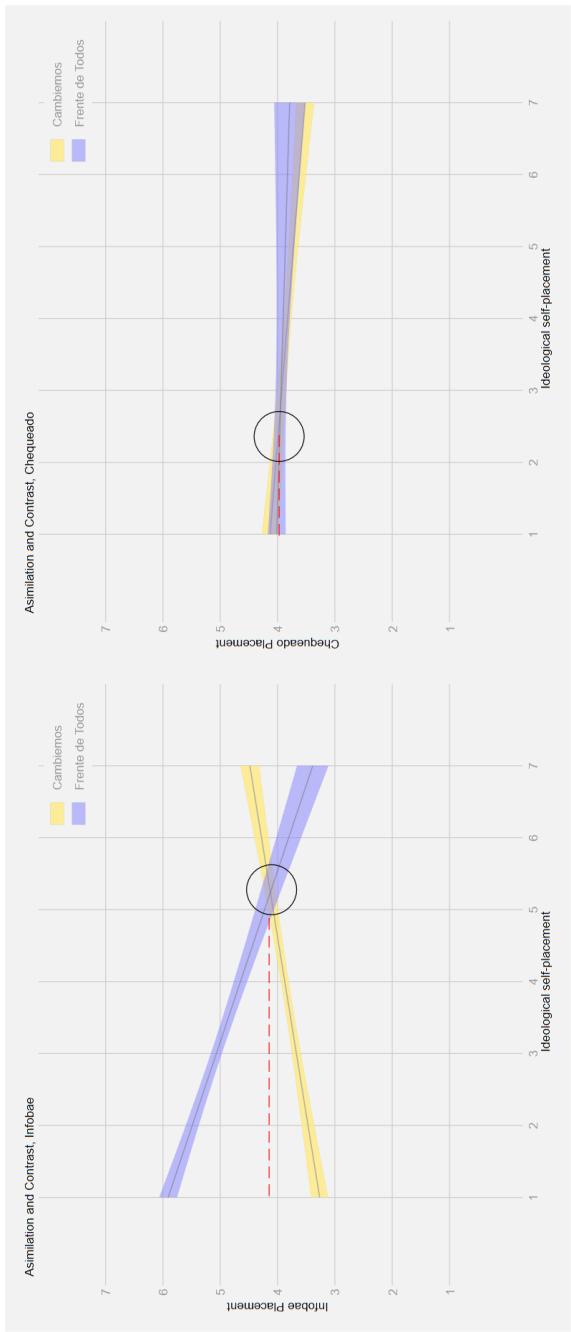
The unbiased reader's perception of the newspaper's position, which we call the newspaper's *unbiased position*, is:

$$y^{\text{unbiased}} := A(x^{\text{unbiased}}) = \alpha_A - \beta_A \frac{\alpha_A - \alpha_C}{\beta_A - \beta_C} = \frac{\beta_A \alpha_C - \beta_C \alpha_A}{\beta_A - \beta_C}.$$

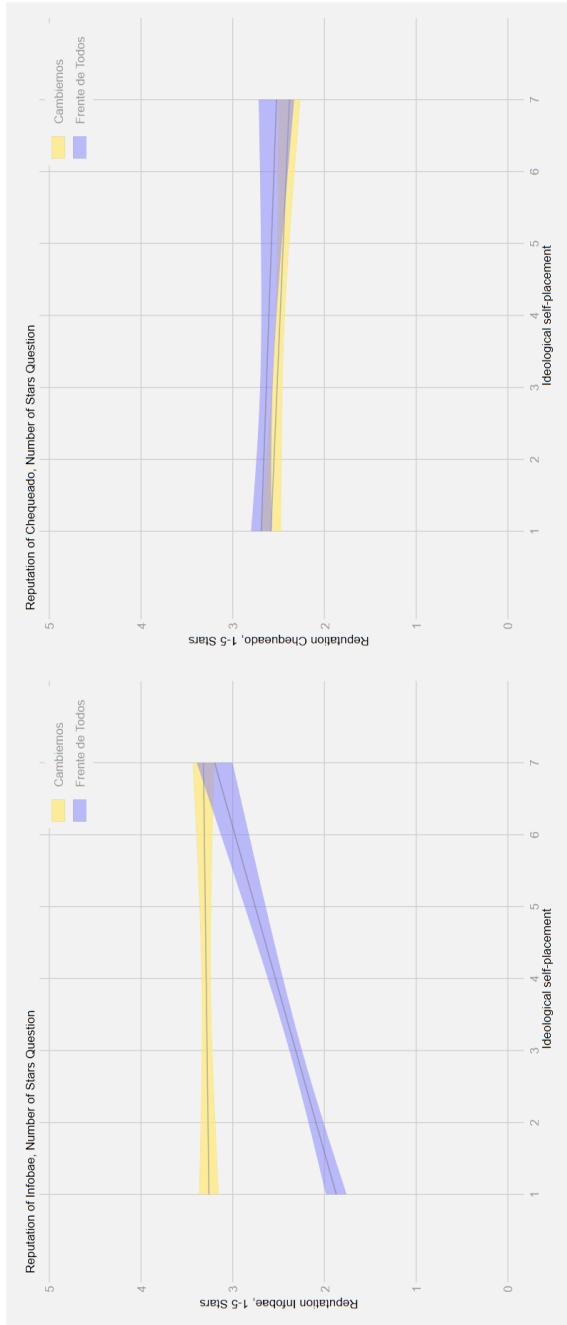
In Figure 1, next to the position of a locally unbiased voter we mark the unbiased placement

of the news organization with a dotted red line. Figure 1 shows two critical features frequent in observational data with news organizations that are assimilated and contrasted: (1) the location at which La Nacion is observed by the locally unbiased voter is modestly conservative while for Pagina|12 is modestly progressive. However (2), readers are not concentrated at the unbiased voter location and they perceive the news organization in widely different places depending on their political leaning. In fact, most progressive voters perceive Pagina|12 as moderate and La Nacion as extreme right while most conservative voters perceive La Nacion as moderate and Pagina|12 as extreme left.

Let us now consider Figure 2, which plots the assimilation and contrast estimates in our survey results for the newspaper Infobae (opposition) and for the fact-checker Chequeado. Readers can readily observe that Infobae is more centrist than La Nación but still significantly affected by assimilation and contrast. The perceived position of Chequeado is very close to the center and, more important, is perceived as less affected by assimilation and contrast. Such centrist placement does not prevent small assimilation and contrast shocks that can be experimentally tested. There is significant sensitivity to pro- and counter-attitudinal treatments, as we will show next.



*Note:* The horizontal axis describes the self-reported location of the respondents. The vertical axis describes the reported location of the conservative newspaper Infobae (left) and the factchecker *Chequeado* (right). Blue lines describe how government supporters of the *Frente de Todos* perceive both organizations and yellow lines describe how supporters of the opposition *Cambiemos* perceive them. Red lines describe the location at which each news organization would be perceived as identical by supporters of both parties. This “unbiased” user location is modestly conservative for Infobae and is close to the middle of the scale (1-7) for *Chequeado*. The contrast line for Infobae is similar to that of La Nación, although both the negative contrast slope is softer and the unbiased location is more centrist. The goal of our experiment is to assess whether pro- and counter-attitudinal adjudications alter the assimilation and contrast slopes of *Chequeado*.



*Note:* The horizontal axis describes the self-reported location of the respondents. The vertical axis describes the reported quality of the conservative newspaper Infobae using a five star rating system (left) and the factchecker *Chequeado* (right). Blue lines describe how government supporters of the *Frente de Todos* rank the organizations and yellow lines describe how supporters of the opposition *Cambiemos* rank them. The reputation of the news organization approximates the valence shock observed in assimilation and contrast. The goal of our experiment is to assess whether pro- and counter-attitudinal adjudications alters the perceived reputation of *Chequeado*.

An independent test of the positive and negative valence shocks after adjudications of TRUE or FALSE come from survey instruments that ask about the perceive quality of a news organization. In our survey, we include a rating question that measures reputation using a five-stars system. Consider Figure 3, which reports the five-star rating of Infobae and Chequeado. As it is possible to observe, the perceived reputation of Infobae is more sensitive to respondents on the left of the political spectruem. Meanwhile, a small negative valence among the opposition supporters is observed in Chequeado, with a small but statistically significant decline in overall reputation independent of ideology.

In the next section we describe the design of our experiment, which analyzes the effect of pro- and counter-attitudinal adjudications on the perceived ideological location and in the reputation of the Chequeado.

## 2 From Theory to Design

Our experiment measures the reputation cost of partisan selective exposure and selective sharing, modeling the response to fact checks that randomly confirm or refute the respondent's beliefs.<sup>5</sup> The reputation costs of factchecking is measured using two different dependent variables: first, we estimate 'assimilation' and 'contrast' models measuring the relationship between self-reported ideological preferences and the respondent reported location of the different news organizations, including Chequeado.<sup>6</sup> Second, we measure if individuals exposed to pro- and counter-attitudinal adjudications report respectively higher or lower "quality marks" for

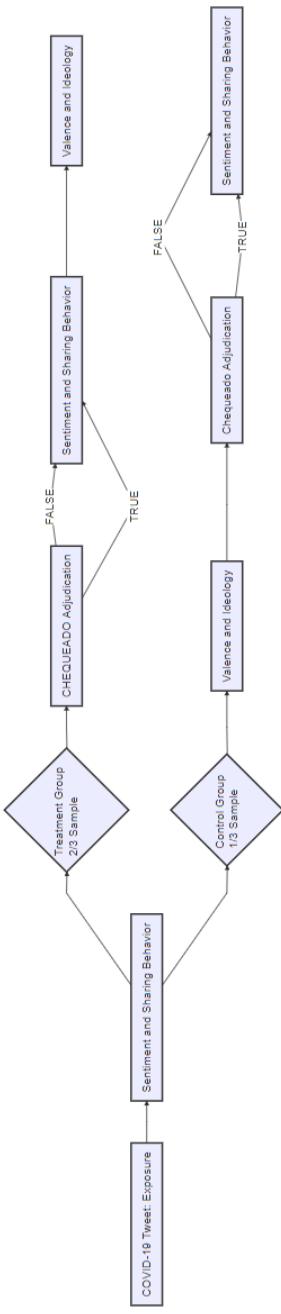
---

<sup>5</sup>It has been well documented that users prefer to spread pro-attitudinal fact-checks over counter-attitudinal fact-checks (Ekstrom and Lai, 2021; Shin and Thorson, 2017). To this end, we measures sharing behavior in the control group, as well as in the TRUE and FALSE adjudications.

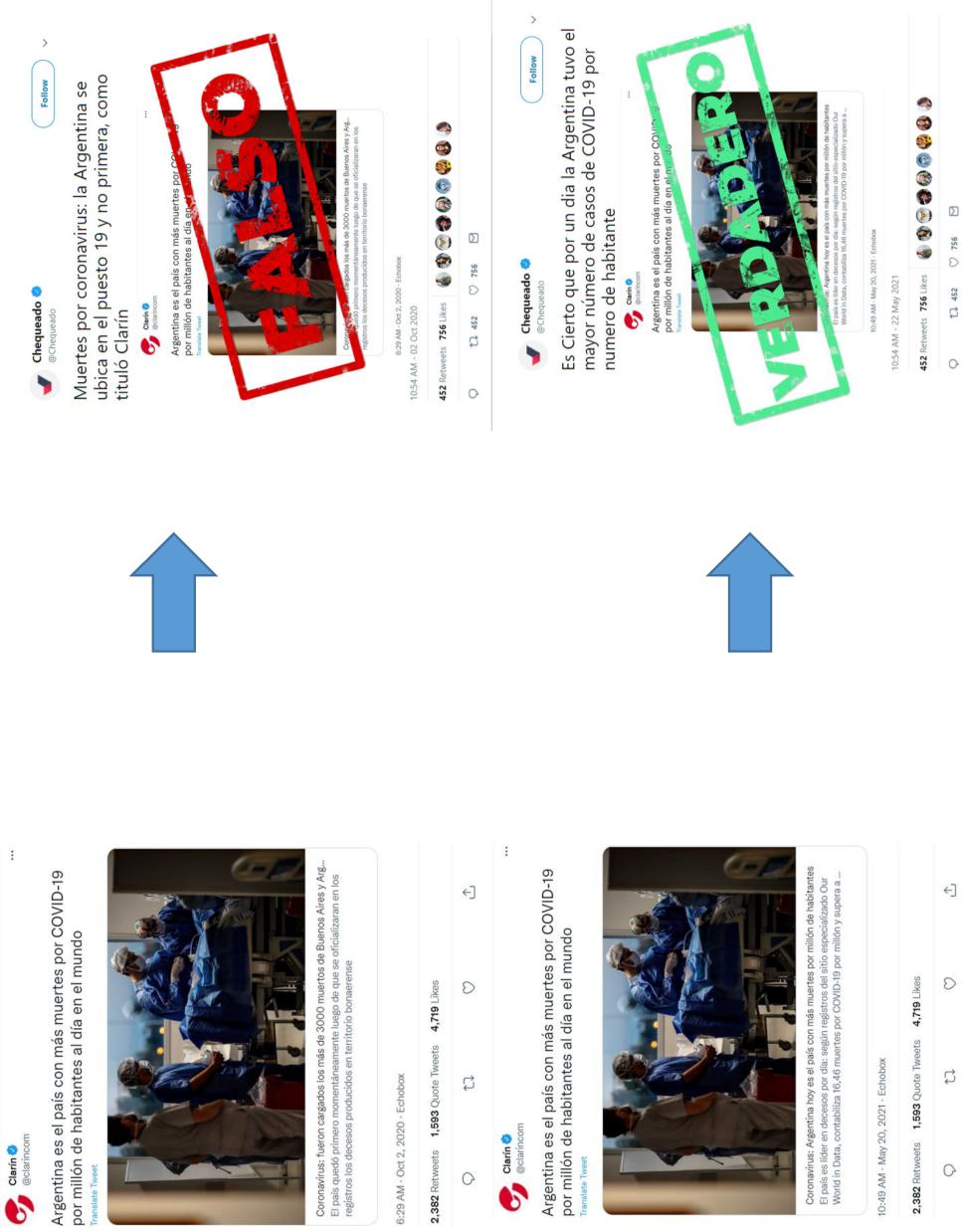
<sup>6</sup>Our design replicates the strategy used in Banks et.al. 2021, who treat respondents to negative Tweet posts by Hillary Clinton and Donald Trump and then measure how negative messages alter the perceived proximity to each candidate. In this case, we model assimilation and contrast effects but in perceived ideological distances to the fact checker.

Chequeado (number of stars to the different news organizations).

Our two-branch two-stage design in Figure 4 exposes respondents to a Tweet that reports the number of COVID-19 deaths in Argentina. We ask respondents if they would share the Tweet and how this Tweet made them feel. We then distract respondents with other questions and split the sample in our treatment and control group. For our treated group, we proceed to adjudicate the initial Tweet was TRUE or FALSE. We ask if they would share the adjudication by the fact checker and whether they believe that the original Tweet was true or false. Most important, half of the individuals are asked questions about the perceived quality and ideological standing of different politicians and news organizations, including Chequeado, while the other half of the respondents is asked these same questions after the fact check. Therefore, the design measures the reputation and perceived ideological location of the fact checker before the intervention and after the intervention. As a result, we will be able to assess whether pro- or counter-attitudinal fact checks alter the perceived quality and ideological position of the fact checker.



**Figure 1** Flow: Respondents observe and react to an original tweet. They are then separated into treatment and control groups. The control group answers questions about ideology and reputation and then is presented with the TRUE or FALSE adjudication. The treated group is presented with the TRUE or FALSE adjudication and then answers the ideology and reputation questions.



**Figure 2** The original tweet in its Clarin Newspaper version is on the left. The TRUE and FALSE adjudications are shown on the right. The TRUE adjudication corresponds to the Tweet of May 20, 2021. The FALSE adjudication corresponds to the Tweet of October 2, 2020.

It is important to highlight that the experiment is not providing misinformation to the respondents. The only deceptions takes place in the rotation of the source organization, which randomly rotates between La Nación, Clarin, Infobae, or Pagina|12. The selected Tweets are TRUE or FALSE on different dates, which ensures that the correction by the fact checker is always accurate. In the TRUE case, the initial Tweet published by Infobae was accurate in May 2021. In the FALSE case, the initial tweet also posted by INFOBAE in November of 2020. This inaccurate information was fact checked three days later. Therefore, no misinformation was presented to respondents but rather a selectively chosen post that was adjudicated as TRUE or FALSE so that it conforms to the actual evidence.

## 2.1 Main Hypotheses

The experimental design yields the following hypothesis that address questions of **social media sharing, adjudication and contrast, and reputation**.

The first hypothesis of our study expects sharing to be more frequent with pro-attitudinal content. Therefore, we expect that respondents that support the government will be less likely to share the Infobae Tweet than respondents that support the opposition. This is consistent with prior studies on network activated frames (Aruguete et al., 2021b), the result of a close relation between cognitive congruence and attention.

$HT_1$ : Pro-attitudinal messages will be shared at a higher rate than counter-attitudinal messages.

The second hypothesis of our study builds on (Aruguete et al., 2021a), showing that adjudications of TRUE are more readily shared than adjudications of FALSE. Therefore, pro-attitudinal confirmation (“it is TRUE that...”) are more likely to be shared than pro-attitudinal refutations (“it is FALSE that...”). This is also true when considering counter-attitudinal confirmations or

refutations. Therefore, all else equal:

The second hypothesis<sup>7</sup> of our study evaluates the effect of pro-attitudinal and counter-attitudinal adjudications on the perceived quality of the fact checker:

*HT<sub>2a</sub>*: Pro-attitudinal confirmations and refutations will increase the perceived quality of the fact checker.

*HT<sub>2b</sub>*: Counter-attitudinal confirmations and refutations will decrease the perceived quality of the fact checker.

The third hypothesis of our study evaluates the consequences of pro-attitudinal and counter-attitudinal adjudications on the perceived ideological distance between the fact checker and the respondent. This hypothesis connects the experiment to the literature on assimilation and contrast, as described in Banks et. al (2021):

*HT<sub>3a</sub>*: Pro-attitudinal confirmations and refutations will decrease the perceived ideological distance between the respondent and the fact checker (assimilation effect).

*HT<sub>3b</sub>*: Counter-attitudinal confirmations and refutations will increase the perceived ideological distance between the respondent and the fact checker (contrast effect).

## 2.2 Dependent Variables

We estimate two different sets of models to test for *HT<sub>2</sub>* and *HT<sub>3</sub>*, one measuring the perceived quality of the fact-checker using a 5-star rating and another one measuring assimilation and contrast effects in perceived ideological proximity.

The dependent variable of the reputation models estimates the number of “5-stars” given by each respondent  $i$  to Chequeado,  $Y_i$ . The variable, described in detail in the supplemental

---

<sup>7</sup>In the original PAP we presented an ancillary hypothesis of an independent positive shock value of the label TRUE compared to the label FALSE. The study of this framing effect is analyzed in a different article, as it is part of a different series of studies on framing.

documentation, ranges from 0 to 5 stars. The dependent variable of the ideological proximity model takes as its dependent variable the distance between the self-reported ideology of the respondent and the reported ideology of “Chequeado”. The variable, described in detail in the supplemental documentation, ranges from 1 (very progressive) to 7 (very conservative). We estimate separate linear specifications for the control group, the pro-attitudinal treatment group and the counter-attitudinal treatment group.

### 2.3 Independent Variables

The first independent of our experiment is a Treatment dummy that takes the value of 1 if the respondents answered the ideology and reputation questions *after* reading Chequeado’s adjudication or a value of 0 if they answer the ideology and reputation questions *before* the adjudication. If the respondents were assigned to the control group, they were still presented with the adjudication later in the survey, allowing us to measure their reactions for validation purposes.

The second independent variable was the TRUE/FALSE assignment, which takes the value of 1 if the respondent is in the group that read Chequeado’s TRUE adjudication and a value of 0 if they are in the group that was assigned the FALSE adjudication.

A third independent variable is a categorical variable that indicates which news organization posted the initial tweet. This variable could take the values of La Nación (conservative-opposition), Infobae (center-right opposition), Clarín (center-opposition), or Pagina|12 (Progressive government supporter). The only difference between each of the posted tweets was the author, while everything else remained identical.

A fourth important covariate was the respondents vote preference for a second round between a Cambiemos ticket (Vidal-Larreta) and a Frente de Todos ticket (Fernandez-Fernandez) if the

general presidential election were to take place next week. The SIF file also presents alternative models using the more general party questions.

A number of other covariates are also considered to validate the robustness of the results, including socio-demographic, partisan, and social media behavior variables. This includes covariates describing the level of engagement of the respondents with the original tweet and with the adjudication as well as time-to-read and time-to-react to the posts. We expect sharing rates to heighten the treatment effects, consistent with Calvo and Ventura (2021). Alternative models are presented in the Supplemental Information File (SIF).

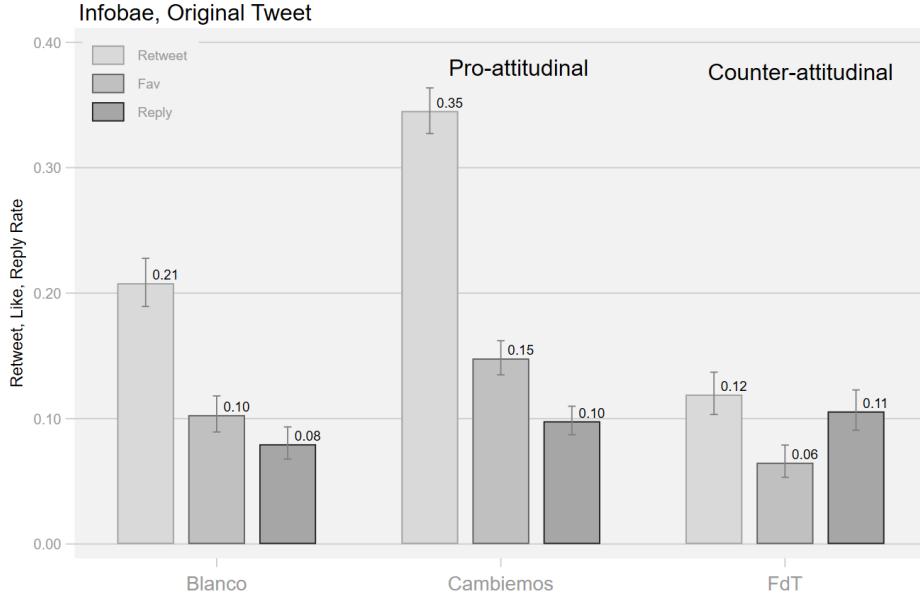
## 3 Results

### 3.1 Sharing

The first hypothesis of this study,  $HT_1$ , expected pro-attitudinal messages to be shared at a higher rate than counter-attitudinal messages. We expected pro-attitudinal sharing of the original tweet by the opposition Cambiemos to be higher than the counter-attitudinal sharing by the supporters of President Alberto Fernandez (Frente de Todos or FdT).

We also expected supporters of Cambiemos to share the TRUE adjudication at higher rates than the supporters of President Fernandez, the FdT. Meanwhile, we expected the the FdT to share the adjudication of FALSE at a higher rate than Cambiemos. Pro- and counter-attitudinal sharing are also crucial preconditions for any testing of the reputation and ideological placement effects. The decision to share content that aligns with the voters preferences is an important marker of the positive and negative content conveyed by the treatments.

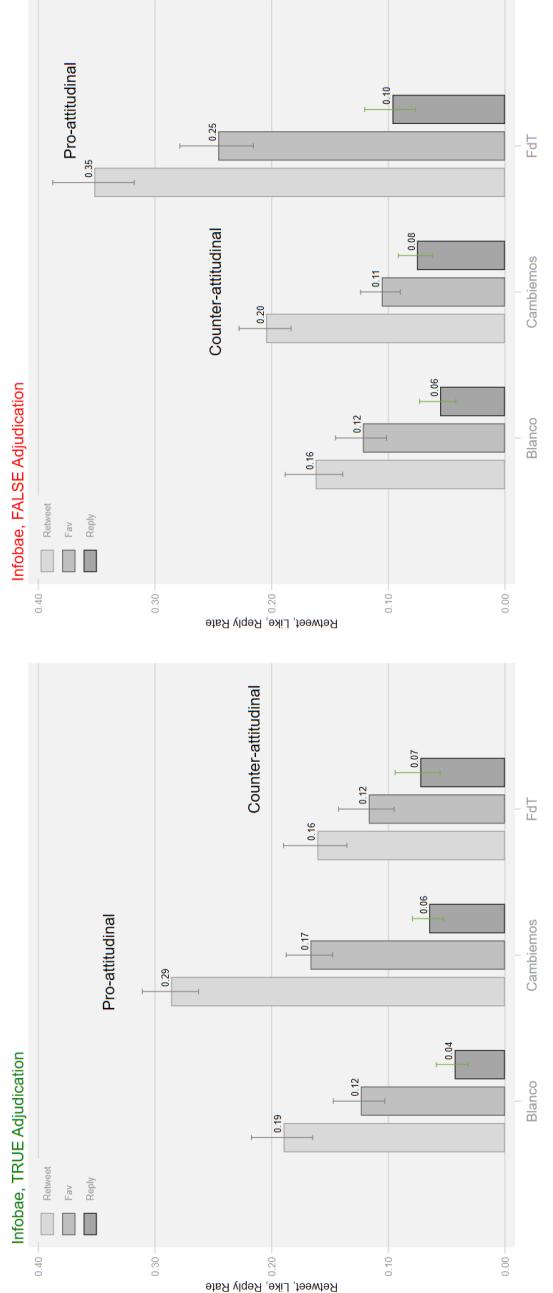
Figure 5 presents results describing the sharing behavior of the respondents to the original tweet. Similarly, Figure 6 presents sharing for the TRUE and FALSE adjudications of



*Note:* The vertical axis describes the share of respondents that indicated their preference to retweet, like, and/or reply the initial tweet of Infobae. Respondents could mark multiple responses. Therefore the sum of the sharing responses may exceed 1.

Chequeado. As it is possible to observe, the pro-attitudinal sharing of the original tweet by Cambiemos supporters is 35%, three times larger than for the FdT voters. The share of likes is almost three times higher, 15% compared to 6%. Only the “reply” rate is similar, consistent with social media evidence showing the injured party to engage in dialogue when treated to cognitively dissonant content, compared to liking or retweeting cognitively congruent content.

The sharing results of the adjudication are equally sound, with Cambiemos supporters sharing the pro-attitudinal TRUE adjudication at almost twice the rate of the FdT supporters. The sharing of the pro-attitudinal adjudication FALSE by the FdT supporters is equally noteworthy, again sharing the Chequeado post 35% of the times compared to 20% by Cambiemos supporters. Overall, results conclusively validate pro-attitudinal sharing of both the initial post and both adjudications by Chequeado.



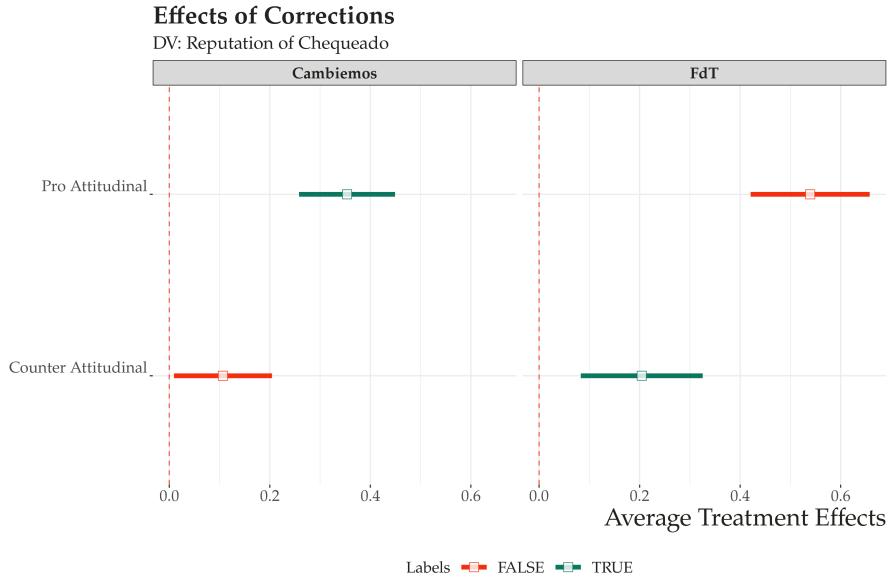
*Note:* The vertical axis describes the share of respondents that indicated their preference to retweet, like, and/or reply the adjudication by Chequeado. Respondents could mark multiple responses. Therefore the sum of the sharing responses may exceed 1. The left plot reports responses to the TRUE adjudication and the right plot reports responses to the FALSE adjudication.

### **3.2 Reputation Effects**

The second hypothesis of this study,  $HT_{2a}$ , stated that pro-attitudinal confirmations and refutations would increase the perceived quality of the fact checker. That is, supporters of Cambiemos treated with a TRUE adjudication and supporters of the FdT treated with a FALSE adjudication would perceive Chequeado as more reputable (more stars). On the other hand, counter-attitudinal confirmations and refutations would decrease the perceived quality of the fact checker.

Figure 7 describes the average treatment effect (by party) when respondents are treated with the TRUE or FALSE adjudications. Green lines indicate adjudications of TRUE and red lines indicate adjudications of FALSE, holding the control group as the baseline. The results of the restricted model that include only the main covariates already provide conclusive evidence of a pro-attitudinal reputation shock. An increase in the number of stars is larger for respondents that support Cambiemos when treated to the adjudication of TRUE and for respondents that support the FdT when treated with an adjudication of FALSE.

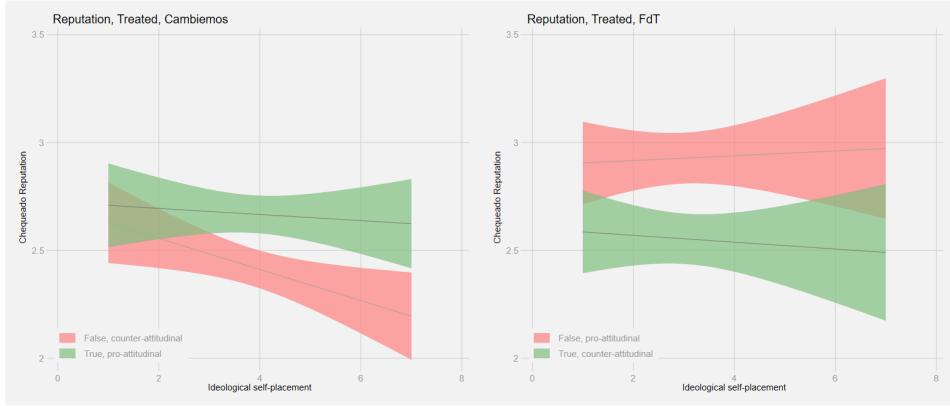
Interestingly, results also show net positive gains vis-á-vis the control group, with respondents reporting higher number of stars even when reading the counter-attitudinal post. Given that no information is provided to the respondents about Chequeado, this increase may reflect an issue-change effect rather than a net gain, as fact checking on COVID-19 may already carry an independent positive charge when compared to fact checking statements by partisans or elected officials. Issue rotation seems an interesting and significant extension to future research that seeks to understand the reputation effects of fact checking.



*Note:* Average treatment effect (ATE) for the restricted reputation model with no controls. Results confirm that pro-attitudinal fact checks increase the reputation of Chequeado compared to counter-attitudinal fact check. However, a positive effect for counter-attitudinal fact check compared to the control group provides evidence of issue-heterogeneity when respondents are untreated. Fact checking information on COVID-19 may provide independent positive valence, likely compared to fact checks of strictly partisan content.

Figure 8 provides a different view of the reputation effects of pro- and counter-attitudinal adjudications, this time conditional on both party and ideology. As it is possible to observe, counter-attitudinal adjudications have a more detrimental effect in the conservative and very conservative subgroup of supporters of Cambiemos. The difference between the pro-attitudinal and counter-attitudinal adjudications for supporters of the FdT, on the other hand, are not affected by the self-reported ideological location of the respondent.

Next we turn our attention to the estimates of assimilation and contrast effects of the TRUE and FALSE adjudications.

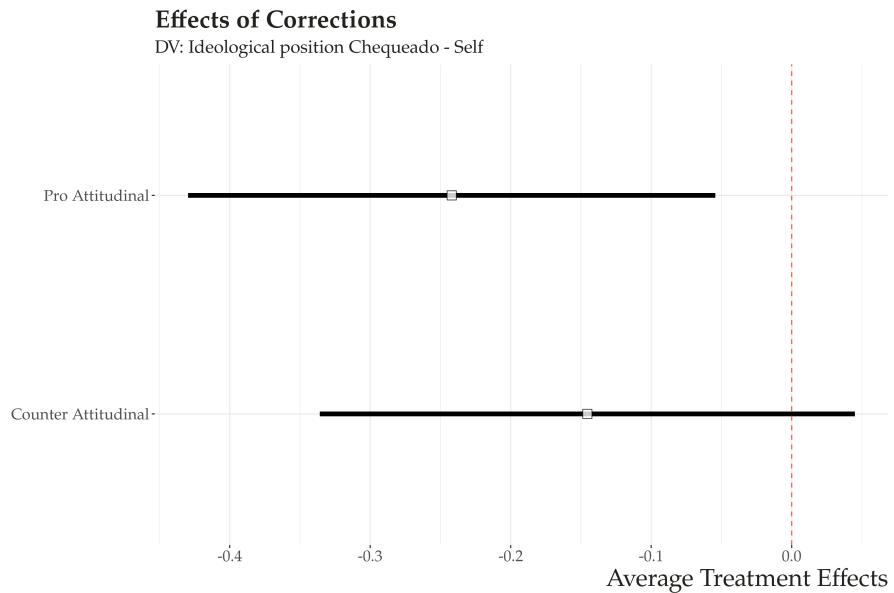


*Note:* Conditional effect of ideology and treatment on reputation. Results provide evidence of a significant decline in reputation for conservative and very conservative supporters of Cambios treated with counter-attitudinal adjudications. The pro- and counter-attitudinal effects on reputation are independent of the self-reported ideological position among the supporters of the FdT.

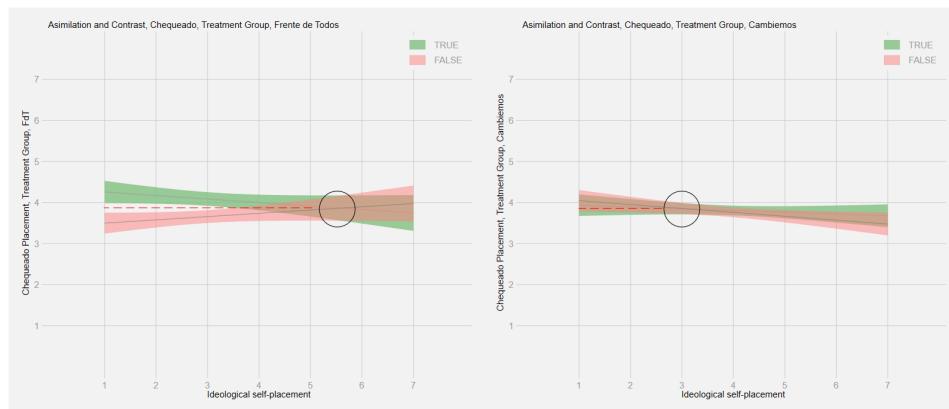
### 3.3 Assimilation and Contrast

The third hypothesis of our study evaluates the consequences of pro-attitudinal and counter-attitudinal adjudications on the perceived ideological distance between the fact checker and the respondent. This hypothesis,  $HT_{3a}$ , expects pro-attitudinal confirmations and refutations to decrease the perceived ideological distance between the respondent and the fact checker (assimilation effect). On the other hand, we expect counter-attitudinal confirmations and refutations to increase the perceived ideological distance between the respondent and the fact checker (contrast effect).

Figures 9 and 10 present the overall results and the assimilation and contrast plots respectively. As expected, we find that pro-attitudinal confirmations and refutations decrease the perceived distance between the respondent and the fact checker. However, the average treatment effect is not positive nor it is statistically significant compared to the control group. Figure 10 clarifies the results, showing that the effect of pro- and counter-attitudinal adjudications have the expected effect on the FdT respondents but not among voters of Cambios.



*Note:* Average treatment effect (ATE) for the directional effect of the treatment on ideological distance, with no controls. Results confirm that pro-attitudinal fact checks reduce the perceived ideological distance compared to counter-attitudinal fact check. However, counter-attitudinal fact check do not produce an aggregate increase in ideological distance.



*Note:* Assimilation and contrast models by party. Results provide evidence of a statistically significant decline in ideological distance after treating FdT respondents to pro-attitudinal adjudications. We also observe a statistically significant increase in distance after counter-attitudinal adjudications. However, results are not statistically significant for treated respondents that support Cambiemos.

## 4 Concluding Remarks

In this paper we implement a survey experiment to measure the reputation cost of publishing pro- and counter-attitudinal fact checks in a polarized political environment. We expect pro-attitudinal fact checks to increase the perceive reputation and reduce the ideological distance reported by readers. Concurrently, we expect counter-attitudinal fact checks to reduce the reputation and increase the perceived ideological distance to the fact checker. Results confirm the expected pro-attitudinal effects on reputation registered in our pre-approved plan (PAP). Results are less robust when measuring the assimilation and contrast effect of an adjudication, with statistically significant results only among supporters of the Frente de Todos. Overall, findings validate that it is possible to moderate the reputation and perceived ideological proximity reported by respondents. Our findings also suggest that further research is needed to clarify the relationship between issue selection and reputation, which is key to understand how prior information about fact checking in the control group.

## References

- Adams, J. F., Merrill III, S., and Grofman, B. (2005). *A unified theory of party competition: A cross-national analysis integrating spatial and behavioral factors*. Cambridge University Press.
- Ahler, D. J. and Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, 80(3):964–981.
- Aldrich, J. H., Schober, G. S., Ley, S., and Fernandez, M. (2018). Incognizance and perceptual deviation: Individual and institutional sources of variation in citizens' perceptions of party placements on the left-right scale. *Political Behavior*, 40(2):415–433.
- Ardèvol-Abreu, A. and Gil de Zúñiga, H. (2017). Effects of editorial media bias perception and media trust on the use of traditional, citizen, and social media news. *Journalism & mass communication quarterly*, 94(3):703–724.
- Aruguete, N., Bachman, I., Calvo, E., Valenzuela, S., and Ventura, T. (2021a). Truth be told: Cognitive and affective moderators of selective sharing of fact-checks on social media.
- Aruguete, N., Calvo, E., and Ventura, T. (2021b). News by popular demand: Ideological congruence, issue salience, and media reputation in news sharing. *The International Journal of Press/Politics*, 4.
- Banks, A., Calvo, E., Karol, D., and Telhami, S. (2021). # polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, 26(3):609–634.
- Bolsen, T., Druckman, J. N., and Cook, F. L. (2014). The influence of partisan motivated reasoning on public opinion. *Political Behavior*, 36(2):235–262.

- Brandtzaeg, P. B. and Følstad, A. (2017). Trust and distrust in online fact-checking services. *Communications of the ACM*, 60(9):65–71.
- Broockman, D. E. and Skovron, C. (2018). Bias in perceptions of public opinion among political elites. *American Political Science Review*, 112(3):542–563.
- Calvo, E., Chang, K., and Dodyk, J. (2021). A model of voting with motivated reasoners: with applications to affect polarization.
- Calvo, E., Chang, K., and Hellwig, T. (2014). Beyond assimilation and contrast: Information effects, ideological magnification, and the vote. *Electoral Studies*, 36:94–106.
- Calvo, E. and Ventura, T. (2021). Will i get covid-19? partisanship, social media frames, and perceptions of health risk in brazil. *Latin American politics and society*, 63(1):1–26.
- Ekstrom, P. D. and Lai, C. K. (2021). The selective communication of political information. *Social Psychological and Personality Science*, 12(5):789–800.
- Granberg, D., Harris, W., and King, M. (1981). Assimilation but little contrast in the 1976 us presidential election. *The Journal of Psychology*, 108(2):241–247.
- Graves, D. (2018). Understanding the promise and limits of automated fact-checking.
- Gunther, A. C. and Schmitt, K. (2004). Mapping boundaries of the hostile media effect. *Journal of Communication*, 54(1):55–70.
- Hansen, G. J. and Kim, H. (2011). Is the media biased against me? a meta-analysis of the hostile media effect research. *Communication Research Reports*, 28(2):169–179.
- Iyengar, S. and Kinder, D. R. (2010). *News that matters: Television and American opinion*. University of Chicago Press.

- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Lee, E.-J. (2012). That's not the way it is: How user-generated comments on the news affect perceived media bias. *Journal of Computer-Mediated Communication*, 18(1):32–45.
- Lee, T.-T. (2005). The liberal media myth revisited: An examination of factors influencing perceptions of media bias. *Journal of Broadcasting & Electronic Media*, 49(1):43–64.
- Lodge, M. and Taber, C. S. (2013). *The rationalizing voter*. Cambridge University Press.
- Mason, L., Wronski, J., and Kane, J. V. (2021). Activating animus: The uniquely social roots of trump support. *American Political Science Review*, pages 1–9.
- Merrill, S., Grofman, B., and Adams, J. (2001). Assimilation and contrast effects in voter projections of party locations: Evidence from norway, france, and the usa. *European Journal of Political Research*, 40(2):199–223.
- Muraoka, T. and Rosas, G. (2021). Does economic inequality drive voters' disagreement about party placement? *American Journal of Political Science*, 65(3):582–597.
- 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.
- Strickland, A. A., Taber, C. S., and Lodge, M. (2011). Motivated reasoning and public opinion. *Journal of health politics, policy and law*, 36(6):935–944.
- Vallone, R. P., Ross, L., and Lepper, M. R. (1985). The hostile media phenomenon: biased perception and perceptions of media bias in coverage of the beirut massacre. *Journal of personality and social psychology*, 49(3):577.

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.