Intentional or inadvertent fake news sharing? Fact-checking warnings and users' interaction with social media content

Alberto Ardèvol-Abreu; Patricia Delponti; Carmen Rodríguez-Wangüemert

Alberto Ardèvol-Abreu ✉
https://orcid.org/0000-0001-8722-5226

Universidad de La Laguna
Dpto. de Psicol. Cognitiva, Social y Org.
Campus de Guajara, Apartado 456.
38200 La Laguna (Santa Cruz de Tenerife),
Spain
aardevol@ull.es

Patricia Delponti
https://orcid.org/0000-0001-9694-867X

Universidad de La Laguna
Dto. de Cien. de la Com. y Trabajo Social
Campus de Guajara, Apartado 456.
38200 La Laguna (Santa Cruz de Tenerife),
Spain
adelpont@ull.edu.es

Carmen Rodríguez-Wangüemert
https://orcid.org/0000-0001-5903-3913

Universidad de La Laguna
Dto. de Cien. de la Com. y Trabajo Social
Campus de Guajara, Apartado 456.
38200 La Laguna (Santa Cruz de Tenerife),
Spain
crodrigu@ull.edu.es

Abstract

The main social media platforms have been implementing strategies to minimize fake news dissemination. These include identifying, labeling, and penalizing –via news feed ranking algorithms– fake publications. Part of the rationale behind this approach is that the negative effects of fake content arise *only* when social media users are deceived. Once debunked, fake posts and news stories should therefore become harmless. Unfortunately, the literature shows that the effects of misinformation are more complex and tend to persist and even backfire after correction. Furthermore, we still do not know much about how social media users evaluate content that has been fact-checked and flagged as false. More worryingly, previous findings suggest that some people may *intentionally* share made up news on social media, although their motivations are not fully explained. To better understand users' interaction with social media content identified or recognized as false, we analyze qualitative and quantitative data from five focus groups and a sub-national online survey (*N* = 350). Findings suggest that the label of 'false news' plays a role –although not necessarily central– in social media users' evaluation of the content and their decision (not) to share it. Some participants showed distrust in fact-checkers and lack of knowledge about the fact-checking process. We also found that fake news sharing is a two-dimensional phenomenon that includes intentional and unintentional behaviors. We discuss some of the reasons why some of social media users may choose to distribute fake news content intentionally.

Keywords

1. Introduction

Social media have become the new 'mass media,' and an important entry point for news content for many citizens (**New-man** *et al.*, 2019). While some of these changes in information consumption habits may have a number of beneficial effects in terms of social capital, political discussion, or civic engagement (**Gil de Zúñiga**; **Jung**; **Valenzuela**, 2012), recent observations have also pointed to the problematic link between social media and fake news (**Allcott**; **Gentzkow**, 2017; **Vosoughi**; **Roy**; **Aral**, 2018). Fake news content distorts informed thinking and comprehension about important issues, with potentially devastating consequences for subjects of social interest such as governance or public health, among others (see **Bode**; **Vraga**, 2018; **Gunther**; **Beck**; **Nisbet**, 2019).

The main social media platforms have recently got down to work against the spread of fake content. Following the 2016 U.S. Election, *Facebook* started using external fact-checkers to identify, label, and reduce the visibility of false publications in the news feeds (**Mena**, 2020). In a similar vein, *Twitter* changed its traditional "hands off position" regarding fake content (**Hern**, 2020b) in the recent context of the coronavirus outbreak. The microblogging company broadened their definition of harmful content to include, among others, denial of health authority recommendations and any kind of "false or misleading information around Covid-19" (**Gadde**; **Derella**, 2020). The company also announced the adoption of a 'manipulated media label' to warn their users about altered videos and photos (*Twitter Safety*, 2020).

> ❝ Fake news content distorts informed thinking and comprehension about important issues, with potentially devastating consequences for subjects of social interest ❞

Fact-checking and labeling fake content in social media may seem a straightforward way to fight misinformation. It seems logical to think that, once warned about the falsehood of a publication, people will stop believing and sharing it. However, while some studies indicate that warning labels may reduce the credibility and the willingness to share the content (**Mena**, 2020), there could also be unintended negative consequences. First, decades-long experimental research in psychology has found that corrections often reduce, but rarely eliminate the continued influence effect of previously encoded misinformation (**Lewandowsky** *et al.*, 2012). Second, and according to the implied truth effect, social media users may wrongly believe that unlabeled false stories in their news feed have been fact-checked and found to be true (see **Pennycook** *et al.*, 2020). Third, the repetition of false information that occurs when fact-checkers correct it (for example, in the common myth versus facts approaches) can paradoxically backfire and increase misbeliefs (see **Cohen** *et al.*, 2007; **Ecker** *et al.*, 2011). Finally, and even more worrisome, a previous study conducted among British *Twitter* users suggest that a significant portion of users may *intentionally* share "exaggerated" (17.1%) and even "made up" (8.9%) news on their social media channels (**Chadwick**; **Vaccari**; **O'Loughlin**, 2018).

In order to better understand the way people interact with social media content once they have been warned about their falsehood and how they use these warnings to decide whether to share it or not, we collected qualitative data from five focus group discussions. Participants were presented with three fact-checked posts, explicitly labeled as 'false' or 'partly false,' and then exchanged their views on the contents and the likelihood to share it with others. A second study, conducted a few days after –at the onset of the Covid-19 lockdown in Spain– collected survey data from a diverse, sub-national convenience sample about respondents' levels of trust in different sources of news, ideological leanings, concern about the coronavirus disease, and intentional fake news sharing behaviors, among others.

2. Fact-checking and labeling of fake content in social media

The main social media sites have routinely portrayed themselves as neutral platforms –not publishers– and bastions of "free expression in speech" (**Marantz**, 2019; **Paul**, 2019). In enacting this role, they been traditionally reluctant to moderate or remove content, especially when it comes to political speech. For example, until very recently, *Facebook* limited their punitive mechanisms to

> "abusive behavior such as account hacking, malware, spam and financial scams" (**Weedon**; **Nuland**; **Stamos**, 2017, p. 3).

Similarly, *Twitter* has been frequently criticized for having too lax policies against misinformation (**Hern**, 2020a). Increasing criticism about the role of social media in spreading fake news (e. g., during the 2016 U.S. election, Brexit referendum, or public health crises such as Zika and coronavirus outbreaks) has progressively moved these platforms toward less neutral policies regarding misinformation. In December 2016, *Facebook* announced a partnership with external fact-checkers –signatories of the *International Fact-Checking Network* code of principles– that would help them verify and, where appropriate, flag fake news stories in the news feed as 'disputed' (**Mosseri**, 2016). One year later, the platform decided to cease using red warning flags next to fake content because, they argued, according to academic research (**Lewandowsky** and colleagues' 2012 piece)

> "strong language or visualizations (like a bright red flag) can backfire and entrench someone's disbelief" (**Smith**; **Jackson**; **Raj**, 2017).

However, in October 2019, and "in order to protect the 2020 U.S. elections", *Facebook* re-revisited their policies against misinformation and started to cover some of the fake content with a grey overlay screen and a 'false' or 'partly false information checked by independent fact-checkers' label (**Rosen** *et al.*, 2019).

Facebook demands country-level fact-checkers to abide by the code of principles of the *International Fact-Checking Network*, which is based on the commitment to nonpartisanship, transparency of sources and funding, and to "open and honest corrections" (*Poynter Institute*, n. d.). Despite the importance of fact-checking for debunking strategies, some previous studies suggest that social media users are, at best, ambivalent about fact checking services, and often openly distrustful of them (**Brandtzaeg**; **Følstad**; **Chaparro-Domínguez**, 2018; **Wenzel**, 2020). In the case of Spain, *Facebook*'s fact-checking partner *Newtral* has been repeatedly –and falsely– accused of colluding with the government to censor and police social media and *WhatsApp* ("*Facebook* califica de falsas," 2020). What is worse, internet users' feelings of pervasiveness of fake news may be connected to a more general informational disbelief (online), which could in turn nurture the idea that the distinction between fact-checked and not-fact-checked information, quality news and fake news, it is more "one of degree rather than a clear [one]" (**Nielsen**; **Graves**, 2017, p. 1; see also **Brandtzaeg**; **Følstad**; **Chaparro-Domínguez**, 2018).

It is therefore necessary to learn more about the way people use these external cues (i.e., false content warnings, partial covering of false posts) to assess the quality and credibility of social media content. Since these debunking strategies online are relatively new and of highly uncertain effectiveness (**Pennycook** *et al.*, 2020; **Smith**; **Jackson**; **Raj**, 2017), we believe it is important to capture and scrutinize social media users' own words and reactions concerning these warnings. This knowledge is necessary to provide more suitable tools to better combat fake news and misinformation. With this goal in mind, we ask our first research question:

> RQ1: How do social media users make use of warning labels and overlays in their assessments of the quality and credibility of online content?

3. Unintentional and intentional fake news sharing in social media

Over the last decade, social media have become a public and private space for citizens "to share, discuss, and contribute to news" and public interest content (**Hermida** *et al.*, 2012). Within this changing context, an increasingly frequent way to become aware of public issues is via suggestions and recommendations from social media contacts (e.g., shared, retweeted, or forwarded content from family members, friends, or coworkers). Social recommendation of news and public affairs content has been proved to be a powerful tool for information dissemination (**An** *et al.*, 2011), but this includes misinformation and disinformation as well. For example, a recent study that examined 126,000 true and false news shared on social media found that false news content

> "spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it" (**Vosoughi**; **Roy**; **Aral**, 2018, p. 6).

In this context, it is important to understand why people may decide or not to share online news content within their social media networks. Based on the uses and gratifications perspective, previous studies have identified a number of social and psychological motivations that may drive news sharing behaviors. These include, among

> The fact of sharing a piece of news content with one's social media contacts involves a certain level of approval of the source and the quality of the information

others, informing and entertaining one's audience, persuading and changing others' opinions (**Chadwick**; **Vaccari**; **O'Loughlin**, 2018; **Wong**; **Burkell**, 2017), or creating shared experiences and a feeling of belonging (**Hermida** *et al.*, 2012; **Ho**; **Dempsey**, 2009). **Lee** and **Ma** (2012) proposed a more comprehensive account that grouped motivations for news sharing into three main groups: socializing (i.e., creating and improving social relationships), status seeking (i.e., being perceived as knowledgeable, credible, or an opinion leader), and information seeking (i.e., creating a 'news archive' that can be retrieved at some point in the future).

From the above outlined theoretical perspectives, the fact of sharing a piece of news content with one's social media contacts involves a certain level of approval of the source and the quality of the information. When an individual shares news-related content, they bring awareness to "something that has potential value" to their contacts (**Wong**; **Burkell**, 2017, p. 2). For this reason, most strategies for minimizing the distribution of fake news have focused on correcting and labeling disinformation. Since most people do not want to be perceived as dishonest or unreliable, it is reasonable to assume that strategies such as labeling fake news or showing 'related articles' that include fact-checkers' refutations will minimize fake news sharing and circulation (see **Rosen**, 2020).

However, some people may still believe fake news content that has been labelled as false by independent fact-checkers. For example, according to the well-examined hostile media perception, partisans and those who are personally involved with an issue tend to perceive 'the media' (in general) and specific news stories (in particular) as biased against their side (**Choi**; **Yang**; **Chang**, 2009; **Giner-Sorolla**; **Chaikem**, 1994). This might also apply to fact-checking organizations, which are occasionally perceived as biased and taking part "in fraud or propaganda" (**Brandtzaeg**; **Følstad**; **Chaparro-Domínguez**, 2018, p. 12). Relatedly, the literature on motivated reasoning has identified several cognitive biases that may influence

how people process fake news and fact-checking labels. Thus, partisan goals may motivate some social media users to actively scrutinize and criticize attitudinally incongruent warnings from fact-checkers and to accept confirming information (e.g., fake news) without questioning it (disconfirmation bias, see **Strickland**; **Taber**; **Lodge**, 2011; **Taber**; **Lodge**, 2006). Another cognitive mechanism refers to individuals' tendency to perceive attitudinally discrepant arguments as weaker and less compelling than congruent ones (prior attitude effect, see **Taber**; **Lodge**, 2006).

More germane to this study, and as pointed out above, people may also have motivations to distribute fake news *intentionally* (i.e., even knowing or at least having the suspicion that the content is false or exaggerated) (**Chadwick** *et al*. 2018). Some previous evidence points to the role of ideological (i.e., advance candidates they favor) and entertaining/trolling motivations to circulate inaccurate stories (see **Allcott**; **Gentzkow**, 2017; **Chadwick**; **Vaccari**; **O'Loughlin**, 2018), but this issue has not been fully explored yet. Social media labels and warnings from fact checkers can provide a valuable opportunity to explore whether and why (some) users share fake content with their contacts. Thus, our final hypothesis and set of research questions are aimed at better understanding reasons and motivations that may lie behind (intentional and unintentional) fake news sharing behavior:

RQ2: How do social media users make use of warning labels and overlays in deciding whether to share content with their contacts?

H: Fake news sharing is a two-dimensional phenomenon that includes unintentional and intentional behaviors.

RQ3: What motivations drive (intentional and unintentional) fake news sharing behavior on social media?

RQ4: What are the main characteristics of social media users that intentionally share fake news with their contacts?

4. Methods

This study uses a mixed method (qualitative-quantitative) approach. We collected data from 1) focus group discussions of social media users, and 2) a survey among residents in the Autonomous Community of Canary Islands (Spain).

Study 1: Focus groups

Five focus group interviews comprised of 31 participants (in groups of 6 to 7) were conducted between March 10 and 12, 2020 –two days before the coronavirus lockdown in Spain. The study was reviewed and approved by the *Ethics Committee* of the *Universidad de La Laguna* (*CEIBA*, registration number 2020-0378). Participants were recruited from a panel of respondents (mean age 38 years) in Tenerife, administered by *Servicios Avanzados de Opinión* (*SAO*), a local public opinion company that was hired for the purpose. Each participant was paid €30 for their time and effort.

In order to reproduce the homogeneous networks through which fake news usually spread (**Allcott**; **Gentzkow**, 2017), participants were grouped by age (*young* and *older* groups) and ideological orientation (*left*, with participants scoring between 1 and 3 in a 1 to 9 left-right ideological scale; *center*, comprised of participants scoring 4 to 6; and *right*, with discussants scoring between 7 and 9). Thus, we generated two left self-identified groups (young, *M* = 30.33 years, and older, *M* = 41.17 years), one center group (without age considerations, *M* = 45.00 years), and two right groups (young *M* = 25.83 years and older *M* = 46.50 years). The discussions were prepared by the researchers and moderated, video-recorded, and transcribed by *SAO*. Sessions lasted between 62 and 90 minutes (*M* = 72 minutes).

Discussions were stimulated by the projection of three *Facebook* posts, partially covered with a grey overlay screen and labeled either as 'false' or 'partly false,' and 'checked by independent fact checkers.' Two of the posts were overtly political, while one was not. Participants were told that the information "had been recently posted on *Facebook*," and included "a notification" (the overlay screen with the afore mentioned labels), without providing further specification. After making sure that the notice was projected long enough to be read, the moderator told participants that it was possible to remove the overlay screen and read the content by clicking on 'see post.'

The moderator then projected the uncovered post and gave participants a paper copy of the text and picture on it. One of the posts (A) falsely accused the Spanish far-right party *Vox* of having voted against an increase in retirement pensions (see fact-check details in: *Maldito bulo*, 2020). Another debunked anti-immigration post (B) claimed that "Romanians, Ecuadorians, and Moroccans eat seven in ten euros of our [Spanish] public healthcare system" (see fact-check in: *Maldita migración*, 2019). Finally, a third non-political post (C) suggested that parcels coming from China could be infected with coronavirus, and remarked that "one gets panicky" when reading things like "the virus lasts up to 12 days on objects… or more" (see report in: *Maldita.es*, 2020). *Facebook* posts were presented in different order to each group, so that ideologically cross-cutting content was discussed in the last place[1]: A-C-B (groups on the ideological left), B-C-A (groups on the right), C-A-B (group on the center).

Study 2: Online survey

Compared to Study 1, the more individual and anonymous nature of an online survey minimizes social desirability bias (particularly regarding undesirable behaviors such as intentional fake news sharing). The questionnaire we used was part of a larger survey about the initial impact of the Covid-19 pandemic in the Autonomous Community of Canary Islands (Spain), conducted between March 28 and April 3, 2020. The same public opinion company (*SAO*) administered the questionnaire using the *LimeSurvey* online tool. *SAO* distributed the survey invitations to their own consumer panel –comprised of residents in the Community above the age of 16– and shared it on their social media. For each completed survey, *SAO* donated €1 to an international NGO to aid fight against Covid-19.

The first block of the questionnaire was designed to estimate the initial impact of the crisis on respondents' learning and professional activities, as well as their assessment of the public and private response to the pandemic. A second and third blocks, of voluntary response, explored respondents' emotional response to the outbreak, their sources of information and specific knowledge about the virus, and their perceptions regarding coronavirus-related fake news and misinformation.

For the third block (news use and fake news perceptions), we obtained valid responses from 471 respondents. This convenience sample was fairly diverse in terms of demographic distribution, but younger and more educated than official population data (2019 Municipal Register of Inhabitants, and Labor Force Survey): *gender* (49.9% women in this study, compared to 50.50% in the official register), *age* (16.77% in the 15-24 age group[2], 23.78% between 25-34, 35.24% between 35-49, 20.59% between 50-64, and 3.61% older than 64 years; compared, respectively, to the official data of 10.53%, 12.94%, 25.89%, 21.48%, and 15.82%), *education* ('1' = less than 5 years of primary education to '10' = doctoral degree; *M* = 7.69, *SD* = 1.66, *Mdn* = 8, undergraduate degree[3]).

5. Results

5.1. Use of warning labels to assess the credibility of social media content

The first research question asked about participants' use of social media warning cues to assess the quality and credibility of the content. Overall, it is striking that focus groups participants barely mentioned *Facebook*'s fact-check overlay. Although the moderator –deliberately– avoided any prompt about fake news or fact-checking, we expected participants to focus their conversations on the visually aggressive overlays including 'false' or 'partly false' labels. But this did not happen. It is as if fact-checkers' evaluations were regarded as just one more thing to consider, but not necessarily a central one. This would not be problematic if participants systematically verified social media content (for example, by comparing information from other media sources). However, this seems to not be the case, as most participants also stated that it is impossible to verify the barrage of information they constantly get via social media. Some of the participants elaborated on the credibility of the presented posts as follows:

> (54 years, right-wing group, about the 'B' post): The thing is that I need more information to analyze this news item […]. I do not have a clear criterion to say whether this is true or not.

> (47 years, right-wing group, about the 'B' post): The data shown here are quite credible […]. For example, we are taking for granted that this is a false news story, but it is not. What happens is that we go too fast, many people just read the headline. But look at it, it is clear, Romanians…

> (46 years, left-wing group, about the 'B' post): It is a piece of news and it was also published in a legal newspaper [the *Facebook* post included a link to a newspaper story], so it must have been verified because that is what the law says. They use sources that cannot be revealed. They cannot reveal their sources to the public, but they do have to reveal them to the court. This is protected by law, and therefore I think [that the information] is not manipulated or incorrect, but only expressed in a very particular way.

In some cases, participants openly cast doubt on the verification process performed by third-party fact-checkers. Although this lack of trust in *Facebook*'s fact-checkers was not generalized among participants (or at least was not explicitly mentioned), it may be understood in the broader context of low trust in social media as a source of news. If an individual does not trust the news they get from social media platforms, one might expect a spillover effect on social media fact-checking (e.g., low trust in social media fact-checkers as well). In Study 2, average trust scores (Table 1) show very low levels for news from social media (*M* = 1.79, *S.D.* = .699) –only above news from instant messaging apps (*M* = 1.66, *S.D.* = .653). In comparison, respondents' levels of trust were higher for traditional media sources such as newspapers (*M* = 2.77, *S.D.* = .609), radio (*M* = 2.71, *S.D.* = .728), and TV news (*M* = 2.61, *S.D.* = .723).

Table 1. Levels of trust in different sources of news (Study 2)

Responses to question: And what level of trust do you have in each of the following media or source?	Mean	Std. dev.	Median
Newspapers (e.g., *El mundo*, *El país*, local newspapers…)	2.77	.609	3
TV news (e.g., *TVE*, *Antena 3*, *TVC*, local TV news…)	2.61	.723	3
Radio news	2.71	.728	3
International media (*BBC*, *The New York Times*…)	2.62	.735	3
Internet searches, blogs, forums	1.99	.765	2
Instant messaging apps (e.g., *WhatsApp*, *Messenger*, *Telegram*…)	1.66	.653	2
Social media (*Facebook*, *Twitter*, *Instagram*, *Youtube*…)	1.79	.699	2
Official information from the government or its agencies	2.89	.880	3
Family and friends	2.45	.724	2

Note. Item non-response and 'don't know' responses ranged from 9.3% (instant messaging apps) to 35.5% (radio news and international media). Four-point Likert scales, from 1 = no trust to 4 = complete trust.

Some participants demonstrated a lack of understanding of what fact-checking is and what its principles are. Others expressed their concern about the verification process and mentioned potential conflicts of interest of the platform or fact-checkers:

> (33 years, right-wing group): When you click on the news article, it says that there is someone fact-checking it. But it is like a Ministry of Truth… I mean, it is like they have the truth. Just like a social networking site has an owner, this one [*Facebook*] has an owner as well, and a political slant. Hasn't it?

The following discussion in the same group (right-wing) illustrates this lack of trust in the warning messages, which seems to be at least partially based on an incomplete understanding of the verification process and the unawareness about the identities of the fact-checkers (i.e., "who is behind that"):

> 1 (27 years): Well, that is great. I did not know about this thing. Because this pops up on *Facebook*, doesn't it? This warning that the information is false, I think it is really good. I did not even know it existed. Because it warns you and then you decide…
>
> 2 (24 years): Yeah, but as the other mate said, you do not know who is behind that…
>
> 1: That is also true. I do not know how reliable that is.
>
> 3 (30 years): I mean, on what basis it is said that this is false? […] Maybe they verify it when it is widely shared, or not. If I write on *Facebook* that I am happy, I do not think that they will check whether that is true or false, you know? […]
>
> 4 (23 years): It is probably a *Facebook* algorithm.
>
> 2: They say they are external… I mean, just below the 'partly or totally false' label, it says that the news has been fact-checked by external staff or something like that. But, what staff, you know? Because they may not even belong to *Facebook* […]. But who is the person behind that?

This suspicion of fact-checkers can get to the point that social media users entirely dismiss the warnings, even without having checked with any other source of information –focus groups participants were not allowed to use their cellphones during the sessions. The following discussion about the 'A' post between participants in the centrist group exemplifies this issue:

> 1 (51 years): I think this news report is characteristic of our time and our politicians.
>
> 2 (62 years): And also true. It is true, it is authentic.
>
> 1: […] I think this news report may be true. It is true that [the decree was passed] in the Parliament with this number of votes in favor and against […]. I think that if there were no such addition [calling some MPs "patriots that are against increasing the retirement pension"], I would have shared it. […]
>
> 2: I think the same as her, exactly the same. [I think this news report] it is true, but it does not explain the reason why they voted against. It seems, at first sight, that the whole world is against [the political party] *Vox* […].

Another participant of the young and right-wing group was also under the impression that the 'B' post was not completely false –despite what *Facebook*'s warnings said. In this case, and based on her explanation, it seems that her assessment of the veracity of the post arises from the consistency between the publication's content and her worldview– a form of motivated reasoning:

> (30 years): I do not think this news report is so false. Perhaps not seven out of ten euros, but it is true that immigration takes a lot of the public expenditure on health… And when you go to the doctor, you either have a private health insurance or they leave you waiting [on the waiting list at the local public hospital], and you are the one who pays taxes, you know? […].

5.2. Fake news sharing as a two-dimensional phenomenon: The role of social media warnings

RQ2 focused on social media users' sharing behaviors *after* they have been given warning signs of the falsehood of the content. More precisely, it inquired about the influence of social media warnings against fake news on participants' content sharing behavior. Data from Study 2 provides some general context for this question and provides empirical support for our hypothesis (fake news sharing as a two-dimensional phenomenon, which includes unintentional and intentional behaviors).

Table 2. Unintentional (a items) and intentional (b items) fake news sharing behaviors

Throughout the coronavirus crisis, how often have you shared news about the disease that…	Never (1)	At some point (values 2-6)
(a) seemed accurate at the time, but you later found was made up or false?	53.7%	46.3%
(a) was exaggerated, and you were not aware of this?	58.8%	41.2%
(b) you suspected was made up when you shared it?	81.4%	18.6%
(b) was exaggerated, and you were aware of this?	83.8%	16.2%
(b) you knew was made up or false at the time?	91.6%	8.4%

Note. Item non-response was 4.2% for each question.

Responses to individual items in Table 2 show that most of our sample never shared fake news content on social media, either intentionally or unintentionally. However, taking 'a' items together, it is noteworthy that 51.2% of the respondents reported having shared "made up or false" or "exaggerated" content during the weeks before the survey without being aware of it at the time. More strikingly, 23.9% of respondents shared "made up or false" or "exaggerated" content online (group of 'b' items) at some point, *and* they were more or less aware of the problem (see **Chadwick** *et al.*, 2018, for comparisons with the British context). A Principal Axis Factoring with Oblimin Rotation (Table 3) indicates two distinct factors accounting for 84.56% of the variance: a 2-item measure of unintentional fake news sharing, and a 3-item measure of intentional fake news sharing, which suits our theoretical expectations.

> Most of the sample never shared fake news content on social media, either intentionally or unintentionally

Table 3. Principal axis factoring of fake news sharing items (direct Oblimin rotation)

News sharing items	Intentional fake news sharing	Unintentional fake news sharing
… seemed accurate at the time, but you later found was made up or false?	-.046	**.940**
… was exaggerated, and you were not aware of this?	.076	**.798**
… you suspected was made up when you shared it?	**.676**	.202
… was exaggerated, and you were aware of this?	**.935**	.021
… you knew was made up or false at the time?	**.870**	-.099

Notes: Values are pattern loadings from the pattern matrix. Primary loadings of a variable on a factor are indicated by boldface type. Kaiser-Meier-Olkin measure of sampling adequacy = .741; Bartlett's test of sphericity, $\chi^2 (10) = 1320.04$, $p < .001$. Both factors were highly correlated ($r = .50$).

Study 1 helped to better interpret the above quantitative findings, as our focus groups discussants commented on about *Facebook* posts that were explicitly labeled as false. Only a few participants explicitly alluded to the 'false information' overlay as the basis for their decision to not share the *Facebook* post. Thus, two participants commented, respectively, on the 'A' and 'C' post as follows:

> (28 years, left-wing group): In principle, I assume what the pop-up window said [to be true]. It said that it was a false news story, and we should not hype it up. On this basis, I would not attach any importance to that news. And I would not share it until it was confirmed by an alternative source or found an echo in the media […].

> (62 years, centrist group): I would not share that because I would be sharing a lie. It is a false news story that even makes me mad, but what you cannot do is spread it. It makes no sense to do it because it is a false story.

5.3. Motivations for (unintentional and intentional) fake news sharing behavior

Some other participants mentioned different reasons why they would interact with (and even share) fake content on their social media (i.e., motivations for fake news sharing, RQ3). For example, some discussants appealed to their freedom of speech and argued that they may share the fake content just because they 'agree' with it "even if the speech is wrong." This assertion fails to distinguish between expressing opinions and providing information about facts, and denotes that the explicit *Facebook*'s warnings did not make much of an impact on them (RQ1):

> (24 years, right-wing group, about the 'B' post): I think this news story may well be shared. I could share, like I could share any other news of any kind. And I think that, you know, there is no problem about that. If, according to their ideas, someone wants to read it, let them. And if they do not want, they will not read it. But I do not think that the way I see the news story will necessarily influence anyone. I share it and that is all. After all, it is my social media account and I would just share it.

> (37 years, left-wing group, about the 'C' post): For me, it is an opinion, and everyone is free to say what they want. Everyone is free to express themselves, whether they are right or not. There is freedom of speech, even if the speech is wrong.

Different from what might have been guessed from the observation of the raw data on purposive false news sharing, some people may not have bad intentions when they share fake content with their contacts. Most of them, however, may be unaware that *Facebook* algorithm interprets their interactions as signals of interest: Sharing or commenting on a fake news post that has not been fact-checked may cause the content to be prioritized and shown higher in the news feed (see **DeVito**, 2017; **Mosseri**, 2018). Some participants explained that they would share the 'C' post to help to debunk it and minimize its circulation (RQ3):

> The analysis identified two distinct factors accounting: a 2-item measure of unintentional fake news sharing, and a 3-item measure of intentional fake news sharing

> (30 years, centrist group): A better way to combat [fake news dissemination] might be the opposite. I mean, do share it, but include your criticism. If we become aware that [the post is getting viral], I think it is a good thing to post a comment or share it on your [*Facebook*] wall and add your critical point of view on it. […] It looks good to

me to reply, express your criticism, and say: "You are lying." Regarding comments, I am known on *Facebook* for replying to everyone. I mean, if I see one of these things, I am going to reply and tell them [what I think].

(28 years, left-wing group): If it was not a viral [post], but something from someone I know, then I would post a comment, as if to say: "You are somewhat mistaken." But it depends on how well I know the other person. If it is content that can go viral, I would not comment on the post. Perhaps I would put a smiley face.

The above direct forms of confrontation (e.g., "you are lying," "you are somewhat mistaken") may not be the best way to correct others. Threatening others' self-integrity could make them reject the correction or even increase their misbeliefs (see **Cohen** *et al*., 2007). A third reason that may explain intentional fake news sharing behavior relates to participants' perception that the truth is provisional, and that something that is found to be false today might be true tomorrow. In the context of the 'C' post, one of the participants argued –not entirely unjustifiably– that in the event of a risk to public health, better to overstate the danger than to understate it:

> Another reason that may explain intentional fake news sharing relates to participants' perception that the truth is provisional, and that something that is found to be false today might be true tomorrow

(46 years, left-wing group): In this case, I would share it. I neither believe nor disbelieve the post [...]. Regardless of whether this news story is true or not, there is a possibility, and that possibility is a true one. Things need to be analyzed objectively, and that possibility is there, and it is true [...] that infected goods may pose a health risk [...]. I am particularly careful with health issues. Regardless of this post being 100% true or not [...], I think it is a good thing to exaggerate a bit when it comes to medical matters, because prevention is better than cure.

5.4. Characteristics of users that intentionally share fake news with their contacts

Finally, RQ4 asked about defining characteristics of social media users that intentionally share fake news. In Study 2, we included items to collect demographic information (gender, age, education level) and measured some relevant attitudes and self-perceptions that may relate to intentional fake news sharing behavior: strength of ideological orientation (regardless of left or right leaning), trust in government information, and concern about Covid-19. In Table 4, we compare those who never shared fake news purposively with those who did (regardless of the frequency).

Table 4. Comparisons of intentional fake news distributors with the rest of the sample

	Intentional fake news distributors	Rest of the sample	Difference tests
Proportion of women	44.40%	52.20%	[a] $\chi 2\,(1) = 1.97$, $p = .16$
Age	39.99 [13.70]	39.34[13.39]	[b] $t(449) = -0.44$, $p = .66$
Education	7.70 [1.62]	7.69 [1.68]	[b] $t(440) = -0.01$, $p = .99$
Strength of ideology	2.50 [1.23]	2.58 [1.20]	[b] $t(419) = 0.65$, $p = .52$
Trust in government info.	2.83 [0.93]	2.91 [0.86]	[b] $t(420) = 0.71$, $p = .48$
Concern about Covid-19	3.35 [0.60]	3.27 [0.69]	[b] $t(449) = -1.18$, $p = .24$

Notes. Intentional fake news distributors, $n = 108$; rest of the sample, $n = 343$. Values shown are group means, except for proportion of women. Square brackets indicate standard deviations. Superscript [a] refers to a chi-square test of independence, and [b] to an independent samples t-test. Education is measured on a ten-point scale (1 = less than 5 years of primary education, to 10 = doctoral degree). Strength of ideological identification is measured on a five-point scale (1 = no ideological identification, to 5 = strong ideological identification, whether with the left or with the right). Concern about Covid-19 is measured on a four-point scale (1 = no concern at all, to 4 = very concerned).

Results in Table 4 show that there are not significant differences between the group of intentional fake news distributors and the rest of the sample –in terms of gender, age, education, strength of ideology, trust in government, or concern about the disease. The proportion of men is higher among intentional fake news distributors (55.6%) compared to non-distributors (47.8%), but the chi-square test did not reach significance: $\chi 2\,(1) = 1.97$, $p = .16$. Fake news distributors also show lower levels of trust in information from the government (2.83 compared to 2.91 among non-distributors), but this difference is again not significant: $t(420) = 0.71$, $p = .48$. Finally, intentional distributors of fake news scored higher on concern about the virus (3.35 among distributors, 3.27 among non-distributors), but the t-test failed to confirm a significant difference between the two groups: $t(449) = -1.18$, $p = .24$. We further tested the effects of our predictors (in Table 4) using OLS regression with heteroskedasticity-consistent standard error estimators (HC0, see **Hayes**; **Cai**, 2007). For this model, we included intentional fake news distribution as the dependent variable, measured in a 1-7 scale (1 = never to 7 = all the time) with a composite index of 'b' items in Table 2 (Cronbach's alpha = .87). However, our set of independent variables did not improve the fit of the model ($R^2 = .012$, $F(6, 385) = 0.38$, $p =. 89$), which is in line with the findings in Table 4.

> There are not significant differences between the group of intentional fake news distributors and the rest of the sample—in terms of gender, age, education, strength of ideology, trust in government, or concern about the disease

6. Discussion and conclusions

This study was aimed to better understand how social media users think and react when exposed to fake content explicitly labeled or recognized as such, considering their own experiences and words. For this, *Facebook*'s overlay screens and 'false' or 'partly false' labels provided a realistic tool for our research design, which is beneficial in terms of ecological validity. *Facebook*'s warnings also facilitated our approach to the thorny and intriguing issue of intentional fake news sharing. Thus, in Study 1 we had the opportunity to evaluate participants' reactions to fake posts without directly having to ask them whether they would share fake news content or not, which is problematic in terms of social desirability. Nonetheless, we directly asked this question in the more personal context of an online survey (Study 2), and our two studies are therefore complementary.

We found that most of our participants in Study 1 did not mention *Facebook*'s warnings and labels during the discussions. This may suggest that most people do not perceive fact-checkers' reports as a major determinant in their decisions about the quality and credibility of the content, but only as one more piece of information to consider. Some focus groups discussants indicated that they "need[ed] more information to analyze th[e] news item," and some others just preferred to follow their predispositions ("the data are quite credible […], it is clear, Romanians"), or misleading hints (a link to a "legal newspaper"). Notwithstanding the above, we need to take participants' lack of references to *Facebook*'s warnings with caution, because most of them also said that they would *not* share the fake posts with their social media contacts. We did not measure the actual magnitude of the influence of fact checkers' warnings on their decision not to share the content, but we need to leave open the possibility that it was an influential but not explicitly declared factor.

Only one participant showed great enthusiasm for the warning labels ("I think it is really good […] because it warns you and then you decide"). Several others expressed suspicion and distrust toward *Facebook*'s fact-checkers ("*Facebook* has an owner, and a political slant") and a critical view of the fact-checking process itself ("It is like a Ministry of Truth"). Some discussants' comments demonstrated an incomplete understanding of fact-checking procedures ("It is probably a *Facebook* algorithm"). These findings may not be exclusive to our focus group participants because they resonate with results from Study 2, and are also in line with **Wenzel**'s (2019) observations in the context of the United States.

According to findings from Study 2, social media and instant messaging apps such as *WhatsApp* are the least trusted news sources (lagging far behind more traditional sources such as newspapers, radio, or TV). If people do not perceive social media as trustworthy news sources, it makes sense that they do not find social media fact-checkers trustworthy either.

> Several discussants expressed suspicion and distrust toward *Facebook*'s fact-checkers and a critical view of the fact-checking process itself

Finally, and as a major contribution of this study, we present some relevant results regarding intentional fake news sharing, following up initial findings from **Chadwick**, **Vaccari** and **O'Loughlin** (2018) in the United Kingdom. This is an almost unstudied problem, with relevant consequences regarding the mechanisms of fake news dissemination –because false content "spreads farther, faster, deeper, and more broadly than the truth because humans, not robots, are more likely to spread it (**Vosoughi**; **Roy**; **Aral**, 2018, p. 6).

Our analysis suggests that fake news sharing is a two-dimensional problem that involves intentional and unintentional behaviors. A strikingly high 23.9% of our sample reported to have had shared "made up or false" or "exaggerated" content about Covid-19 intentionally at some point during the weeks before the survey. We did not find differences between intentional fake news distributors and the rest of the sample in terms of demographic characteristics and other tentative predictors: strength of political identification, trust in government, or concern about the disease. This is in line with previous findings from **Chadwick**, **Vaccari** and **O'Loughlin** (2018) about 'dysfunctional news sharing,' although their study did not distinguish between intentional and unintentional sharing of misinformation (their dependent variable includes both types of behaviors). Future research should explore specific personality traits that may better predict intentional fake news sharing. For example, Machiavellianism –briefly defined as a manipulative personality– may positively predict these democratically dysfunctional behaviors. From a theoretical point of view, it makes sense that manipulative individuals show a reduced commitment to the truth (see **Geis**; **Moon**, 1981), because their manipulative interests –for example, to make others vote in a certain way– may be placed first. Conversely, those scoring high on agreeableness may feel more concerned about deceiving others and may therefore refrain from sharing content if they are not certain of its accuracy.

Findings from Study 1 shed some light on the reasons behind this 'democratically dysfunctional' behavior. And not all motives for intentional fake news sharing are evil, as one may first imagine. Thus, some participants stated that they would share fake posts to "include [thei]r criticism" or to let others know that the story is false. This could however have the opposite effect to that intended, because social media algorithms may interpret this interaction as a signal of interest and prioritize the content in the news feed. Also, motivated skepticism and backfire

> Not all motives for intentional fake news sharing are evil, as one may first imagine

effects –particularly for political content– may render these correction attempts counterproductive (see **Ecker** *et al.*, 2011; **Taber**; **Lodge**, 2006). The latter is even more likely if the wording of corrections threatens the self-integrity of the person who originally posted the content (with expressions such as "you are lying" or

> Some participants explained that they would share the fake posts to help to debunk them and minimize their circulation

"you are somewhat mistaken"). Other participants emphasized the possible social utility of sharing certain alarmist messages (the 'C' post) in the context of a public health risk, "regardless of whether [it] is true or not", "because prevention is better than cure."

Other motivations for intentionally sharing fake content connect more with a perceived difficulty to disentangle opinions from facts in political and public interest matters. Thus, some discussants mentioned that the freedom of speech should entitle everyone to share negatively fact-checked content –such as the 'C' post–. Other participant argued that he would share another political post because he agreed with the way the story was presented and, ultimately, it was his social media account. For these participants, some of the presented *Facebook* posts represent "an opinion," which could be "right or not" but should not be censored. While it is true that this may be a reasonable position in some specific situations, it is also true that one of the dangers of a "post-truth era" is that "people [may] stop believing in facts altogether" (**Lewandowsky**; **Ecker**; **Cook**, 2017, p. 355). If there is nothing right or wrong, if all is debatable and knowledge is perceived as elitist, modern societies may constitute a breeding ground for fake news, misinformation, and conspiracy theories (**Lewandowsky**; **Ecker**; **Cook**, 2017).

One of the limitations of our empirical approach may be a certain vulnerability to social desirability bias. Exploring the issue of intentional fake news sharing in a group context (i.e., our focus group design) can be very useful because it reproduces the environments in which misinformation usually circulates. However, we cannot ignore that most people do not like to be seen as liars or manipulators. It is therefore likely that the less virtuous motivations for fake news sharing have remained unreported and undetected in our study. Further research should employ alternative designs to minimize the magnitude of this bias. Interestingly, a relatively large proportion of respondents of our online survey acknowledged that they had intentionally shared fake news content about Covid-19, which suggest that the anonymity of this research tool made it easier for them to report this socially undesirable behavior. This makes us confident that our mixed methods design was appropriate.

Based on the above findings, we offer five recommendations to educators, journalists, fact-checking organizations, and social media companies to strengthen the positive effect of fact checking-efforts and minimize impact of false news.

- Efforts should be made to better explain to the public the common principles and practices of fact-checking (for example, those contained in Poynter Institute, n.d.). It should be clear to everyone that fact-checking is not a form of censorship and does not deal with opinions but with facts, nor is (yet) an algorithm-centered process.

- Fact-checkers should be perceived as neutral and non-partisan. In countries such as Spain and the United States, some of the most popular fact-checking organizations have being accused of being unfair and biased (*Newtral*, 2020; **Walker**; **Gottfried**, 2019). As a prerequisite for building trust, people need to know that fact-checking sources are neutral and not at the service of political or economic interests.

- In connection with the above points, people should be better informed about the selection criteria fact-checkers use for choosing content to be verified. Social media users should be actively encouraged to send information to be fact-checked, so that they will feel that they are part of the process (see, for example, *Maldito bulo*, 2018).

- Social media platforms should better explain that interacting with false content –even to criticize or correct it– can boost the post's placement in the news feed. Users should be aware of this weakness of social media algorithms and minimize any interaction with fake posts.

- Finally, any measure intended to address the 'human factor' in the dissemination of fake news should be aware that fake news sharing is a bidimensional problem, which includes intentional and unintentional behaviors. These different behaviors will probably require different mitigation measures and strategies.

> Some discussants appealed to their freedom of speech and argued that they may share the fake content just because they 'agree' with it "even if the speech is wrong"

7. Notes

1. The group on the ideological center was considered to be on the center-left. Out of the seven participants attending this focus group, five of them positioned themselves slightly on the left pole of the scale (a score of 4), while two of them considered themselves to be in the exact political center (a score of 5).

2. The youngest respondents in our study were 17 years old.

3. According to data from the last *Labor Force Survey* (fourth quarter of 2019), 27.9% of population in the Canary Islands completed high education; 51.8% finished some level of secondary education; and 13.8% has only primary education.

See complete information in:
https://www.ine.es/jaxiT3/Tabla.htm?t=6369

8. References

Allcott, Hunt; **Gentzkow, Matthew** (2017). "Social media and fake news in the 2016 election". *Journal of economic perspectives*, v. 31, n. 2, pp. 211-36.
https://doi.org/10.1257/jep.31.2.211

An, Jisun; **Cha, Meeyoung**; **Gummadi, Krishna**; **Crowcroft, Jon** (2011). "Media landscape in *Twitter*: a world of new conventions and political diversity". In: *Proceedings of the 5th International AAAI Conference on weblogs and social media* (*ICWSM*), pp.18-25.
https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2825/3283

Bode, Leticia; **Vraga, Emily K.** (2018). "See something, say something: correction of global health misinformation on social media". *Health communication*, v. 33, n. 9, pp. 1131-1140.
https://doi.org/10.1080/10410236.2017.1331312

Brandtzaeg, Petter-Bae; **Følstad, Asbjørn**; **Chaparro-Domínguez, María-Ángeles** (2018). "How journalists and social media users perceive online fact-checking and verification services". *Journalism practice*, v. 12, n. 9, pp. 1109-1129.
https://doi.org/10.1080/17512786.2017.1363657

Chadwick, Andrew; **Vaccari, Cristian**; **O'Loughlin, Ben** (2018). "Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing". *New media & society*, v. 20, n. 11, pp. 4255-4274.
https://doi.org/10.1177/1461444818769689

Choi, Jounghwa; **Yang, Myengja**; **Chang, Jeongheon J. C.** (2009). "Elaboration of the hostile media phenomenon: the roles of involvement, media skepticism, congruency of perceived media influence, and perceived opinion climate". *Communication research*, v. 36, n. 1, pp. 54-75.
https://doi.org/10.1177/0093650208326462

Cohen, Geoffrey L.; **Sherman, David K.**; **Bastardi, Anthony**; **Hsu, Lilian**; **McGoey, Michelle**; **Ross, Lee** (2007). "Bridging the partisan divide: self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation". *Journal of personality and social psychology*, v. 93, n. 3, pp. 415-430.
https://doi.org/10.1037/0022-3514.93.3.415

DeVito, Michael A. (2017). "From editors to algorithms: a values-based approach to understanding story selection in the Facebook news feed". *Digital journalism*, v. 5, n. 6, pp. 753-773.
https://doi.org/10.1080/21670811.2016.1178592

Ecker, Ullrich K. H.; **Lewandowsky, Stephan**; **Swire, Briony**; **Chang, Darren** (2011). "Correcting false information in memory: manipulating the strength of misinformation encoding and its retraction". *Psychonomic bulletin & review*, v. 18, n. 3, pp. 570-578.
https://doi.org/10.3758/s13423-011-0065-1

Gadde, Vijaya; **Derella, Matt** (2020). "An update on our continuity strategy during Covid-19". *Twitter blog*, April 1.
https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html

Geis, Florence L.; **Moon, Tae H.** (1981). "Machiavellianism and deception". *Journal of personality and social psychology*, v. 41, n. 4 pp. 766-775.
https://doi.org/10.1037/0022-3514.41.4.766

Gil de Zúñiga, Homero; **Jung, Nakwon**; **Valenzuela, Sebastián** (2012). "Social media use for news and individuals' social capital, civic engagement and political participation". *Journal of computer-mediated communication*, v. 17, n. 3, pp. 319-336.
https://doi.org/10.1111/j.1083-6101.2012.01574.x

Giner-Sorolla, Roger; **Chaiken, Shelly** (1994). "The causes of hostile media judgments". *Journal of experimental social psychology*, v. 30, n. 2, pp. 165-180.
https://doi.org/10.1006/jesp.1994.1008

Gunther, Richard; **Beck, Paul, A.**; **Nisbet, Erik C.** (2019). "Fake news and the defection of 2012 Obama voters in the 2016 presidential election". *Electoral studies*, v. 61.
https://doi.org/10.1016/j.electstud.2019.03.006

Hayes, Andrew F.; **Cai, Li** (2007). "Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation". *Behavior research methods*, v. 39, n. 4, pp. 709-722.
https://doi.org/10.3758/BF03192961

Hermida, Alfred; **Fletcher, Fred**; **Korell, Darryl**; **Logan, Donna** (2012). "Share, like, recommend". *Journalism studies*, v. 13, n. 5-6, pp. 815-824.
https://doi.org/10.1080/1461670X.2012.664430

Hern, Alex (2020a). "Twitter to remove harmful fake news about coronavirus". *The guardian*, March 19.
https://www.theguardian.com/world/2020/mar/19/twitter-to-remove-harmful-fake-news-about-coronavirus

Hern, Alex (2020b). "Fake coronavirus tweets spread as other sites take harder stance". *The guardian*, March 4.
https://www.theguardian.com/world/2020/mar/04/fake-coronavirus-tweets-spread-as-other-sites-take-harder-stance

Ho, Jason Y. C.; **Dempsey, Melany** (2010). "Viral marketing: motivations to forward online content". *Journal of business research*, v. 63, n. 9-10, pp. 1000-1006.
http://doi.org/10.1016/j.jbusres.2008.08.010

Lee, Chei-Sian; **Ma, Long** (2012). "News sharing in social media: the effect of gratifications and prior experience". *Computers in human behavior*, v. 28, n. 2, pp. 331-339.
https://doi.org/10.1016/j.chb.2011.10.002

Lewandowsky, Stephan; **Ecker, Ullrich K. H.**; **Cook, John** (2017). "Beyond misinformation: understanding and coping with the 'post-truth' era". *Journal of applied research in memory and cognition*, v. 6, n. 4, pp. 353-369.
https://doi.org/10.1016/j.jarmac.2017.07.008

Lewandowsky, Stephan; **Ecker, Ullrich K. H.**; **Seifert, Colleen M.**; **Schwarz, Norbert**; **Cook, John** (2012). "Misinformation and its correction: continued influence and successful debiasing". *Psychological science in the public interest*, v. 13, n. 3, pp. 106-131.
https://doi.org/10.1177/1529100612451018

Maldita.es (2020). "Coronavirus y paquetes procedentes de China: ¿cuánto dura el virus vivo en objetos?". *Maldita.es*, 12 abril.
https://maldita.es/maldita-te-explica/2020/03/12/coronavirus-y-paquetes-procedentes-de-china

Maldito bulo (2020). "No, ni Vox ha votado en contra de subir las pensiones y los salarios mínimos ni ha sido el único partido que se ha abstenido: BNG y la CUP también se abstuvieron". *Maldito bulo*, 6 febrero.
https://maldita.es/malditobulo/2020/02/06/no-ni-vox-ha-votado-en-contra-de-subir-la-pensiones-y-los-salarios-minimos-ni-ha-sido-el-unico-partido-que-se-ha-abstenido-bng-y-la-cup-tambien-se-abstuvieron

Maldito bulo (2018). "Tres novedades en pruebas: servicio de *WhatsApp*, *Bulobot* de *Facebook* y verificador de imágenes". *Maldito bulo*, Julio 18.
https://maldita.es/malditobulo/2018/07/18/3-novedades-en-pruebas-servicio-de-whatsapp-bulobot-de-facebook-y-verificador-de-imagenes

Maldita migración (2019). "No: rumanos, ecuatorianos y marroquíes no consumen 7 de cada 10 euros de la sanidad pública". *Maldita migración*, 28 junio.
https://migracion.maldita.es/bulo/no-rumanos-ecuatorianos-y-marroquies-no-consumen-7-de-cada-10-euros-de-la-sanidad-publica

Marantz, Andrew (2019). "Facebook and the 'free speech' excuse". *The New Yorker*. October 31.
https://www.newyorker.com/news/daily-comment/facebook-and-the-free-speech-excuse

Mena, Paul (2020). "Cleaning up social media: the effect of warning labels on likelihood of sharing false news on *Facebook*". *Policy & internet*, v. 12, n. 2, pp. 165-183.
https://doi.org/10.1002/poi3.214

Mosseri, Adam (2019). "Bringing people closer together". *Facebook newsroom*, January 11.
https://about.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together

Mosseri, Adam (2016). "Addressing hoaxes and fake news". *Facebook newsroom*, December 15.
https://about.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-new

Newman, Nic; **Fletcher, Richard**; **Kalogeropoulos, Anthonis**; **Nielsen, Rasmus-Kleis** (2019). *Reuters Institute digital news report 2019.* UK: Reuters Institute for the study of Journalism. University of Oxford. ISBN: 978 1 907384 61 5.
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf

Newtral (2020). "Facebook califica de falsas las afirmaciones que vinculan el límite del reenvío en *WhatsApp* a los equipos de verificación". *Newtral*, 13 abril.
https://www.newtral.es/facebook-limite-reenvio-whatsapp/20200413

Nielsen, Rasmus-Kleis; **Graves, Lucas** (2017). *News you don't believe: Audience perspectives on fake news.* UK: Reuters Institute for the study of Journalism. University of Oxford.
https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FINAL_download.pdf

     12

Paul, Kari (2019). "Zuckerberg defends *Facebook* as bastion of 'free expression' in speech". *The guardian*, October 17.
https://www.theguardian.com/technology/2019/oct/17/mark-zuckerberg-facebook-free-expression-speech

Pennycook, Gordon; **Bear, Adam**; **Collins, Evan T.**; **Rand, David G.** (2020). "The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings". *Management science* (ahead of print version), pp. 1-14.
https://doi.org/10.1287/mnsc.2019.3478

Poynter Institute (n. d). *International fact-checking network fact-checkers' code of principles*.
https://www.poynter.org/ifcn-fact-checkers-code-of-principles

Rosen, Guy (2020). "An update on our work to keep people informed and limit misinformation about Covid-19". *Facebook newsroom*, April 16.
https://about.fb.com/news/2020/04/covid-19-misinfo-update

Rosen, Guy; **Harbath, Katie**; **Gleicher, Nathaniel**; **Leathern, Rob** (2019). "Helping to protect the 2020 US elections". *Facebook newsroom*, October 21.
https://about.fb.com/news/2019/10/update-on-election-integrity-efforts

Smith, Jeff; **Jackson, Grace**; **Raj, Seetha** (2018). "Designing against misinformation". *Medium*, December 20.
https://medium.com/facebook-design/designing-against-misinformation-e5846b3aa1e2

Strickland, April A.; **Taber, Charles S.**; **Lodge, Milton** (2011). "Motivated reasoning and public opinion". *Journal of health politics, policy and law*, v. 36, n. 6, pp. 935-944.
https://doi.org/10.1215/03616878-1460524

Taber, Charles S.; **Lodge, Milton** (2006). "Motivated skepticism in the evaluation of political beliefs". *American journal of political science*, v. 50, n. 3, pp. 755-769.
https://doi.org/10.1111/j.1540-5907.2006.00214.x

Twitter Safety (2020). "We know that some tweets include manipulated photos or videos […]". Twitter, April 20.
https://twitter.com/TwitterSafety/status/1224799838895607809

Vosoughi, Soroush; **Roy, Deb**; **Aral, Sinan** (2018). "The spread of true and false news online". *Science*, v. 359, n. 6380, pp. 1146-1151.
https://doi.org/10.1126/science.aap9559

Walker, Mason; **Gottfried, Jeffrey** (2019). "Republicans far more likely than Democrats to say fact-checkers tend to favor one side". *Pew Research Center*, June 27.
https://www.pewresearch.org/fact-tank/2019/06/27/republicans-far-more-likely-than-democrats-to-say-fact-checkers-tend-to-favor-one-side

Weedon, Jen; **Nuland, William**; **Stamos, Alex** (2017). *Information operations and Facebook. White paper*, Facebook, April 27.
https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf

Wenzel, Andrea (2019). "To verify or to disengage: Coping with 'fake news' and ambiguity". *International journal of communication*, v. 13, pp. 1977-1995.
https://ijoc.org/index.php/ijoc/article/view/10025

Wong, Lorraine Y. C.; **Burkell, Jacquelyn** (2017). "Motivations for sharing news on social media". In: *Proceedings of the 8th International conference on social media & society*, pp. 1-5.
https://doi.org/10.1145/3097286.3097343