

SHARING FACT CHECKING CORRECTIONS IN POLARIZED POLITICAL ENVIRONMENTS: A STUDY OF CONTEXT AND DISAMBIGUATION

NATALIA ARUGUETE (*Corresponding author*)
Conicet & Universidad Nacional de Quilmes. Buenos
Aires. Argentina
natalia.aruguete@unq.edu.ar
nataliaaruguete@gmail.com
ORCID id: <https://orcid.org/0000-0002-1571-9224>

ERNESTO CALVO
Universidad de Maryland. Maryland. Estados Unidos.
ecalvo@umd.edu
<https://orcid.org/0000-0001-5770-1391>

ABSTRACT

Social media platforms have long been considered a source of polarization. They are also considered a key mechanism for amplifying misinformation and spreading false, distorted, and decontextualized information. In such polarized environments, fact-checking interventions are of the utmost importance. This article aims to understand whether polarizing partisan messages also hinder the circulation of fact-checking corrections. Specifically, we test whether exposure to polarizing political messages next to fact-checking messages alters the propensity of users to share messages (destructive interference) or increases their circulation (constructive interference). Using a survey experiment conducted in Argentina during the National Midterm Election of 2022, we measure the propensity to share fact checks when placed alongside partisan messages. Our results show that polarizing political messages increase the propensity to share fact checks (constructive interference) instead of reducing circulation. This is a positive outcome, demonstrating that interventions to reduce the spread of misinformation are more likely to be shared in partisan political environments. Therefore, while polarization increases the rate at which misinformation is produced, it may also facilitate the circulation of its corrections.

KEYWORDS

Affective polarization, fact checking, social media, Argentina, Disambiguation

COMPARTIR CORRECCIONES DE VERIFICACIÓN DE HECHOS EN ENTORNOS POLÍTICOS POLARIZADOS: UN ESTUDIO SOBRE CONTEXTO Y DESAMBIGUACIÓN.

Como citar este artículo / Citation: Aruguete, Natalia y Ernesto Calvo (2024). Sharing fact checking corrections in polarized political environments: A study of context and disambiguation, *Revista Internacional de Sociología* 82 (4): e264. <https://doi.org/10.3989/ris.2024.82.4.1305>

Copyright: © 2024 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia de uso y distribución Creative Commons Reconocimiento 4.0 Internacional (CC BY 4.0)

Recibido: 30/11/2023. Aceptado: 19/07/2024.

Publicado: 04/02/2025.

RESUMEN

Las redes sociales son señaladas frecuentemente como una fuente potencial de polarización. Actualmente, también son un vector significativo para la circulación de desinformación y la propagación de información falsa, distorsionada y descontextualizada. Nuestro objetivo en este artículo es entender si este contexto de polarización también dificulta la circulación de correcciones de verificación de hechos. Específicamente, nos preguntamos si la presencia simultánea de mensajes políticos polarizantes y verificaciones de hechos reduce la propensión a compartir mensajes (interferencia destructiva) o aumenta su circulación (interferencia constructiva o impulso). A través de un experimento realizado en la Encuesta Nacional Electoral Argentina 2021/2022, analizamos la preferencia por compartir verificaciones de hechos cuando se observan junto a mensajes políticos. Nuestros resultados muestran que la presencia de mensajes políticos polarizantes aumenta la propensión a compartir verificaciones de hechos. Este es un resultado positivo, demostrando que las intervenciones destinadas a reducir la circulación de contenido falso pueden tener éxito incluso en contextos de alta polarización.

PALABRAS CLAVE

Polarización afectiva, verificación de hechos, redes sociales, Argentina, Desambiguación.

INTRODUCTION

Digital social media messages serve as a catalyst that energizes users and mobilizes social collectives. They may be used to challenge the power-that-be and democratize public debate, as in the Arab Spring and the BLM movements, among many international protest events. However, the digital social media environment may also heighten undemocratic discourses, polarize voters, and reinforce existing political, economic, and cultural hierarchies. Such is the dual nature of social media platforms in the current digital environment. They increase unregulated political exchanges among users while simultaneously heightening the prevalence of cognitively congruent messages, uncivil discourse, and misinformation. During elections, positive expressions of voter preferences coexist with the negative reinforcement of partisan speech. This is a context in which misinformation messages thrive. This is also the environment in which fact-checkers' interventions are most needed.

Our goal to experimentally evaluate the relationship between partisan attitudes, misinformation, and fact-checking interventions. Social media platforms have long been considered a source of polarization. They are also considered a key mechanism for amplifying misinformation and spreading false, distorted, and decontextualized information. In such polarized environments, fact-checking interventions are of the utmost importance. This article aims to understand whether polarizing partisan messages hinder the circulation of fact-checking corrections. Specifically, we test whether exposure to polarizing political messages next to fact-checking messages alters the propensity of users to share messages (destructive interference) or increases their circulation (constructive interference). More generally, the experiment in this article provides a window into the effect of context in social media interactions. We show that the interpretation and salience of a social media message changes when it is paired next to other messages, and evaluate the effect of context in fact-checking interventions.

To study the relationship between partisanship, misinformation, and fact-checking, we consider the case of Argentina, a country with high levels of ideological, political, and affective polarization. As noted by Facundo Cruz, Argentina has experienced the emergence of a "polarized bi-coalitionism", resulting from the convergence of "institutional incentives and strategic decisions of the main party leaders" (Cruz, 2021, p. 116). The 2011 enactment of Open, Simultaneous, and Compulsory Primaries (PASO, in Spanish) played a pivotal role in heightening the competitiveness of the electoral cycles, anchored in two major coalitions that encompassed a large majority of the national electorate (Cruz, 2021).¹ Since 2011,

the ideological debate has become more intense, with center-left and center-right coalitions advancing policy agendas broadly perceived by voters as distinct. Similar to the United States, partisan identities have become crucial for mobilizing voters and tagging out-group partisans as an existential threat to democracy, thereby portraying elections as a fight that defines the soul of the nation (Iyengar et al., 2019; Schuliaquer & Vommaro, 2020). With heightened polarization, "anger politics" has become a fertile area of research, a frequent response to out-group discourses. That is, an affective "response to the electoral challenges that (the opponent) implies". (Mason, 2015, p. 14)

Our interest lies in understanding the effect of contextualized messages in partisan environments. We consider two user behaviors frequently observed in social media in Argentina's highly polarized scenario: (1) the propensity to share congruent social media messages that reinforce in-group identities; and, (2) the effect of context on users' sharing decisions, considering how paired messages compete for the users' attention. In our experiments, two messages are placed side by side, competing for the user's attention. Paired messages alter the salience and interpretation of the messages' content, which is described in the literature as "disambiguation" (Gilbert, 2006). That is, a process by which context removes uncertainty and changes the meaning of a message. We aim to describe the effect of context as a process of disambiguation that alters the salience and interpretation of social media messages.

To measure the effect of context on fact-checking interventions, we measure the intent to share (message amplification) when a fact-check is placed next to either a second fact-check or a partisan message. A message, defined as the *primary*, is affected by the presence of another message, an interfering message or *secondary*, which can be cognitively similar/dissimilar and ideologically congruent/incongruent. While the primary message is held constant, the secondary message varies. As a result, the importance and interpretation of the primary message is altered by the presence of the secondary one. The experiment, conducted after the second wave of the Argentine General Survey (between February 9 and 16, 2022), also aims to discern the types of effects elicited by messages among the most partisan users, politically sophisticated individuals, and those displaying heightened interest in issues directly impacting the political group they identify with.

ter-right coalition Juntos por el Cambio (JxC), led by Mauricio Macri and Gabriela Michetti, secured a victory in the 2015 presidential elections. After four years in the presidency, and two years of economic crisis, Mauricio Macri's government failed to secure a second term against the opposition, led by Alberto Fernández and Cristina Fernández.

1 Following three governments led by a center-left variant of the Peronism (*Frente para la Victoria*), the cen-

The article is organized as follows. In the first section, we review the current literature on affective polarization and content activation in social media, with particular attention to the partisan and cognitive incentives to share information. We define social media activation as the act of making content available to the users' peer network through actions such as "liking", "sharing", "quoting" or "commenting". In this particular experiment, we focus on the act of "sharing", which is the simplest social media behavior that makes content available between our online peers. In the second section, we describe how changes to the context (secondary message) affect the propensity to share the primary message. We introduce readers to the main hypotheses of our experiment and discuss the experimental design. In the third section, we present our empirical findings, showing how the same content has different probabilities of being shared depending on the content of co-occurrent social media posts. Finally, we discuss the limitations of our study and some future venues for extending this work.

LITERATURE

Affective Polarization

A prevalent interpretation of polarization posits that our fellow citizens hold political preferences that are radically different from ours. However, an alternative perspective is that polarization increases not because our policy preferences differ but because we are increasingly motivated to *like* our in-group peer network and *hate* out-group users. This is generally described as *affective polarization*. Even if our policy preferences remain unchanged, we perceive larger distances to members of the other parties as the intensity of our affective reactions, connections to peers, and friendship linkages grow. The affective polarization hypothesis contends that perceived polarization increases as politics becomes more relevant and our affective connections to others intensify. This type of polarization, *affective polarization*, is one of the most studied political communication phenomena in the last ten years. (Aruguete & Calvo, 2023)

Since the 1980s, the study of party identities and polarization has been characterized by the coexistence of two distinctive approaches. The instrumental approach to polarization understands increasing distances among voters due to how ideologies and policy agendas diverge across party supporters, predicated on substantive policy outcomes. "Real policy" explains oppositional stances, such as a preference for more stringent or less stringent abortion laws, more generous or less generous social benefits, etc. According to this paradigm, party identity aggregates collective interests, with ideology serving as an information shortcut that summarizes the preferences of citizens (Fiorina, 1981; Franklin & Jackson, 1983). Conversely, the expressive approach posits that the

stability of party affiliations remains resilient to short-term economic and political fluctuations, occasionally transcending preferences for specific issues and the inherent power dynamics of political parties. In this dual framework, Huddy and her colleagues advocate for a complementary integration of both approaches to elucidate individuals' positioning in response to significant events. (Huddy, Mason, & Aarøe, 2015)

The salience of the expressive approach has given rise to a recent scholarly trajectory denominated as "affective polarization," which emphasizes the affective distance declared by voters towards opposing political parties when exposed to political messages (Webster & Abramowitz, 2017; Mason, 2015, 2016). Affective polarization, as Iyengar and Westwood (2015) posit, encapsulates the proclivity of individuals aligning with a specific political party to harbor negative perceptions of supporters belonging to the out-group—that which an individual does not affiliate with—and, conversely, to view members of the in-group positively. This antipathy towards the out-group has intensified, even when citizens express moderate disparities in their perceptions of public policies. Notably, contemporary divisions and schisms are no longer confined to the conventional left-to-right ideological spectrum, thereby challenging the instrumental perspective on politics, given that users' affective responses transcend rational interpretations of events.

The considerations individuals incorporate when defending the standing of the political space they identify with and when opposing adversaries surpass the confines of a rational understanding of the political landscape and extend to voter behavior vis-à-vis political party decisions. Mason (2015) contends that, while positions on public matters influence behavior, they do not surpass the determinative effects of identity alignments. Partisan emotions, according to Mason (2016), are triggered by actors or political messages perceived as threatening. Far from a clear demarcation between affective and political polarization, Druckman and colleagues advocate systematically exploring the relationship between emotions and individuals' positions on political issues. (Druckman et al., 2021)

McCoy and collaborators characterize polarization as a process whereby the myriad differences within a society align in a singular dimension. Consequently, individuals perceive politics and society through an "Us versus Them" dichotomy (McCoy, Rahman, & Somer, 2018). Similarly, the concept of "sorting" elucidates a realignment of diverse attributes in public life, wherein political-party divisions align with social, religious, racial, and other disparities (Mason, 2016).

[polarization] occurs when these differences become aligned within (normally two) camps with mutually exclusive identities and interests (Lozada, 2014; Somer, 2001, 2016a). It is the alignment of opinions

under a single identity, rather than the radicalization of opinion, that 'crystalizes interests into opposite factions' and threatens stability. (McCoy et al., 2018, p. 18)

The definition of affective polarization by Iyengar and Westwood (2015) underscores the propensity of individuals supporting a specific political party to harbor negative sentiments toward the out-group members and positive sentiments toward their in-group peers. These negative sentiments towards members of the opposing parties often intensify through exposure to messages received in their daily interactions at the workplace, in family dialogues, and more relevant here, when reading and interacting on social media platforms. (Barberá et al., 2015)

SELECTIVE ATTENTION AND NEWS SHARING ON SOCIAL MEDIA

Social media have recurrently been identified as a potential locus of polarization, primarily attributable to their adeptness in customizing content, delineating user communities based on browsing proclivities, and triggering the proliferation of content cascades. In the contemporary milieu, they function as a consequential vector for disseminating misinformation, promulgating spurious, distorted, decontextualized information. This prompts an inquiry into the dynamics assumed by polarization within the digital ecosystem. Within social media networks, certain messages undergo extensive dissemination and robust sharing, while others languish in obscurity, failing to capture the public's attention. In this contextual domain, production costs notably diminish, and content consumption precipitates its widespread propagation.

The phenomenon of selective exposure to messages on social media becomes manifest as individuals actively seek information that aligns cognitively with their preferences and preconceived beliefs. The role played by interpretive frames of political messages on social media networks is of paramount significance, which can activate identity-driven partisan responses, thereby shaping the extent to which an appeal necessitates a factual response vis-à-vis an act of political alignment.

Interactions among on-line users within these networks do not adhere to a hierarchical logic; rather, they conform to a networked pattern of content circulation, influenced by both individual motivations and structural ramifications. Decisions and reactions of users dynamically alter the frequency of words, hashtags, and links encountered on social media. At the subjective level, the proximity between users emerges from individual inclinations to integrate into homogeneous structures and affiliate with communities sharing common values. This selective exposure implies that users disseminate information that aligns cognitively with their preferences. By establishing connections with ideologically aligned peers, users

expose themselves to posts that reinforce their existing assumptions. The resultant "homophilic association," coupled with the proclivity to share cognitively congruent messages, engenders topological consequences that condition the circulation of information and, by extension, the behavior of users. The formation of communities on social media is thus explicated through the interplay of both individual user behavior (cognitive congruence and selective attention) and the inherent organic logic of the network's topology. (Aruguete, 2021)

The extant literature on news sharing (Aruguete et al., 2023; Kumpel et al., 2015), the salience of issues (Petrocik, 1996; Kaplan et al., 2006; Epstein and Segal, 2000; Bélanger and Meguid, 2008), and the activation of network frames (Aruguete et al., 2023) substantiates the assertion that messages propelled by influencers and the media's coverage of events can significantly alter the perceived significance of topics among individuals. The importance attributed to these matters can undergo transformation even when individuals' positions on a particular issue remain unaltered. (Kahneman, 2011)

In the digital ecosystem, the contents we encounter never materialize in isolation from their contextual milieu. Boczkowski, Mitchelstein, and Matassi (2018) proffer the concept of "incidental news consumption" to elucidate that news permeates our awareness during moments of leisure, implying that we assimilate information within a context characterized by varying levels of attention and the coexistence of disparate information types. For instance, Alberto Fernández's tweet critiquing Mauricio Macri, a focal point of this study, will elicit distinct emotional responses contingent upon the situational context in which we encounter the post and the associated contents.

Numerous mechanisms exist through which the contents we peruse undergo alteration within the informational context. The influence of context, particularly the editorial processes shaping our interpretation of a message, has undergone a marked escalation within the current multi-platform ecosystem, where we are incessantly subjected to heightened information levels. This constitutes the primary concern of our scholarly endeavor, where we seek to reassess the intricate relationship between message senders and receivers in political communication, recognizing that instances wherein we encounter a message "in isolation" from its context are indeed scarce.

FACT-CHECKING IN AN ERA OF POLARIZATION

In recent years, the dissemination of misinformation—comprising spurious information devoid of political intent, as opposed to information wielded as a weapon to inflict harm—has increasingly challenged democracies, increasing their vulnerability to political shocks and decreasing institutional resilience.

The phenomenon of misinformation strategies is not isolated; rather, it finds increased traction in polarized scenarios. In some countries, influential political actors have assumed the role of primary promoters of false content with a destabilizing purpose. In other societies, responses to misinformation circulated in social media and digital platforms have been consensual, cross-cutting, and collective.

Political and media institutions have long grappled with elevated levels of civic distrust, attributable in part to the sheer quantity and speed of information production and dissemination. This undermines the genuine possibility of fostering a reflective outlook and allocating physical time for content verification. Social media, nonetheless, have emerged as a central vector in the public crisis triggered by the dissemination of false news regarding Covid-19, exerting irreversible effects on the already eroded trust in institutions. The convergence of distortions and deliberate lies on clinical matters becomes an explosive combination when the societal demand for greater certainties about the virus incentivizes the proliferation of publications with incomplete, erroneous, or false information. Interventions by fact-checkers were often supported by the majority of political actors when addressing clinical issues, such as vaccine effectiveness. However, these corrections faced resistance in some cases and reverted to their pre-pandemic forms when misinformation became intertwined with political or health responsibilities. The perception that a fact-checker is not ideologically biased, and that their corrections are technically accurate, is what we define as reputation. Users' trust in the professional activity of the fact-checker depends on this reputation. However, reputational capital can be affected when users read that their beliefs or the beliefs of supporting political actors are refuted in a correction.

SELECTIVE SHARING OF FACT-CHECKS

The more we inhabit social platforms, the greater the probability of being exposed to the content that circulates there. Few social media users seek political information, and even fewer seek fact-checks, but they are still affected by what they inadvertently come across online.

What factors regulate the propensity to share fact-checks among social media users based on how they process these corrections? The literature has shown that people prefer to spread fact-checks that support their attitudes more than those that challenge their positions. This is due to two factors. First, denials demand a greater cognitive load than affirmations. "In social media, a higher cognitive burden will result in disengagement of so-called system 1 processing—that is, fast, automatic, and affective response". (Kahneman, 2011, cited in Aruguete et al. 2023) Second, confirmations carry positive valence compared to refutations. Thus, users report not only a greater

intention to share pro-attitudinal confirmations (rated as "true") than pro-attitudinal refutations (rated as "false") but also a greater propensity to retweet counter-attitudinal confirmations than counter-attitudinal refutations.

Although studies addressing the motivations to share fact-checks of false information are not abundant, previous research suggests that this behavior is motivated, in part, by partisan goals. In this article, our objective is to understand whether a context of polarization can hinder the circulation of fact-checking corrections. Every message is read and interpreted in its context. In this case, our question specifically aims to determine whether the presence of polarizing political messages paired with fact-checking posts reduces the propensity to share messages or increases their circulation.

HYPOTHESIS

The primary hypothesis of our investigation posits that partisans will exhibit a heightened motivation to disseminate content across social networks. This proposition aligns with antecedent research on network-activated frames, elucidating a proximate correlation between cognitive congruence, attentional dynamics, and political dedication.

Hypothesis 1: Partisans and anti-partisans will display higher sharing rates than non-partisans and independents.

The second hypothesis contends that the sharing propensity will be more pronounced among users exposed to pro-attitudinal content, denoting messages that fortify their initial perspectives on a given event. Advocates are expected to disseminate content in alignment with the prevailing values within their in-group at a heightened frequency relative to content circulating in the out-group.

Hypothesis 2: Partisans will display higher sharing rates for posts by in-group candidates and lower sharing rates for out-group candidates.

The third hypothesis intricately models the contextual ramifications of posts mutually shared among partisan users. This dynamic is projected to amplify the sharing frequency of messages issued by the in-group authority (i.e., the candidate with whom the user identifies) and diminish the inclination to share messages issued by the candidate representing the party with which the user disagrees. The author of an original tweet is defined as an "authority," while the user who shares information from another is defined as a "follower." Consequently, we anticipate that supporters of Alberto Fernández will augment their sharing rates for messages congruent with their party affiliation when juxtaposed with posts from the opposing candidate, Mauricio Macri. Conversely, they are expected to temper their inclination to share messages concomitantly paired with corrections from the Argentine fact-checking organization, Chequeado.

Hypothesis 3a: Users will display higher sharing rates of fact-checking posts when messages are convergent (constructive interference).

Hypothesis 3b: Users will display lower sharing rates of fact-checking posts when messages are divergent (destructive interference).

METHOD

To examine these hypotheses, we developed and implemented a survey experiment in the second wave of the Argentine General Survey 2022. The Institutional Review Board (IRB) granted human subjects' approval for the Argentine survey on February 3, 2022 [1825785-3]. Overall project approval was registered under the identification code [1825785-1] "COVID-19, Trust, and Misinformation."

A national representative sample was recruited from Netquest's Behavior Panel, which collects the digital trace data with registered consent from users. Separate survey consent was requested at the beginning of the survey, and a disclaimer provided respondents with information on how to contact the researchers or IRB if needed. The experiment was included in a nationally representative online survey, encompassing a cohort of 2,420 respondents from Argentina. This initiative constituted the second wave of a panel conducted from February 9 to February 16, 2022, with data collection facilitated by Netquest. The survey instrument unfolded in two distinct stages. The first stage involved standard inquiries concerning party identification, voting preferences, and political knowledge, which contributed to developing a political sophistication indicator. Additionally, identity-related questions were posed to gauge the significance of variables such as race, ethnicity, religion, and gender.

The second stage featured a survey experiment employing paired tweet combinations designed for various treatments. In a survey experiment, each "treatment" describes a subgroup that receives specific information different from what is provided to another. The goal is to compare the effect of different framings on the interpretation of both groups of respondents. After exposure to these paired tweets, respondents were prompted to answer a series of questions about the tweet they were "more likely to share," the one they were "more likely NOT to share," and the one they "expected to be commented on in their preferred news outlet." Questions (1) and (2) were formulated to account for the cascade activation process (Aruguete, Calvo, & Ventura, 2023; Aruguete & Calvo, 2018), while question (3) aimed to quantify biases in selective attention within the respondent's environment. As we have already mentioned, selective attention on social media indicates that we pay attention to some users and topics to the detriment of other users we could follow and countless issues with which we could engage. (See Figure 2)

This study's primary objective seeks to analyze the disambiguation process occurring when a message is consumed alongside other messages. As illustrated in Figure 1, the instrument design comprised twelve (12) analyses resulting from six possible combinations of tweet pairs, encompassing four variations of posts: two political posts and two fact-checks. The overarching goal was to scrutinize the preference for one of the two messages and measure the amplification rate of respondents in response to changes in context, specifically when the contents accompanying the original message underwent alterations. The treatments encompassed four types of posts: a post from the former Argentine president, Alberto Fernández (2019-2023); a post from the former president Mauricio Macri (2015-2019); a Chequeado post correcting misinformation about the spread of Covid-19; and a Chequeado correction addressing misinformation related to masks used to prevent virus transmission during the Covid-19 pandemic.

We derived a model with a dependent variable "sharing" that takes the value of 1 if the respondent shared a tweet and 0 if it did not. Our results include two noteworthy measures: the average marginal effects and the marginal means (Hainmueller et al., 2014; Leeper et al., 2020). The former provides insights into our primary choice of sharing for average effects, while the latter is employed to identify effects within subgroups in our pairing analyses.

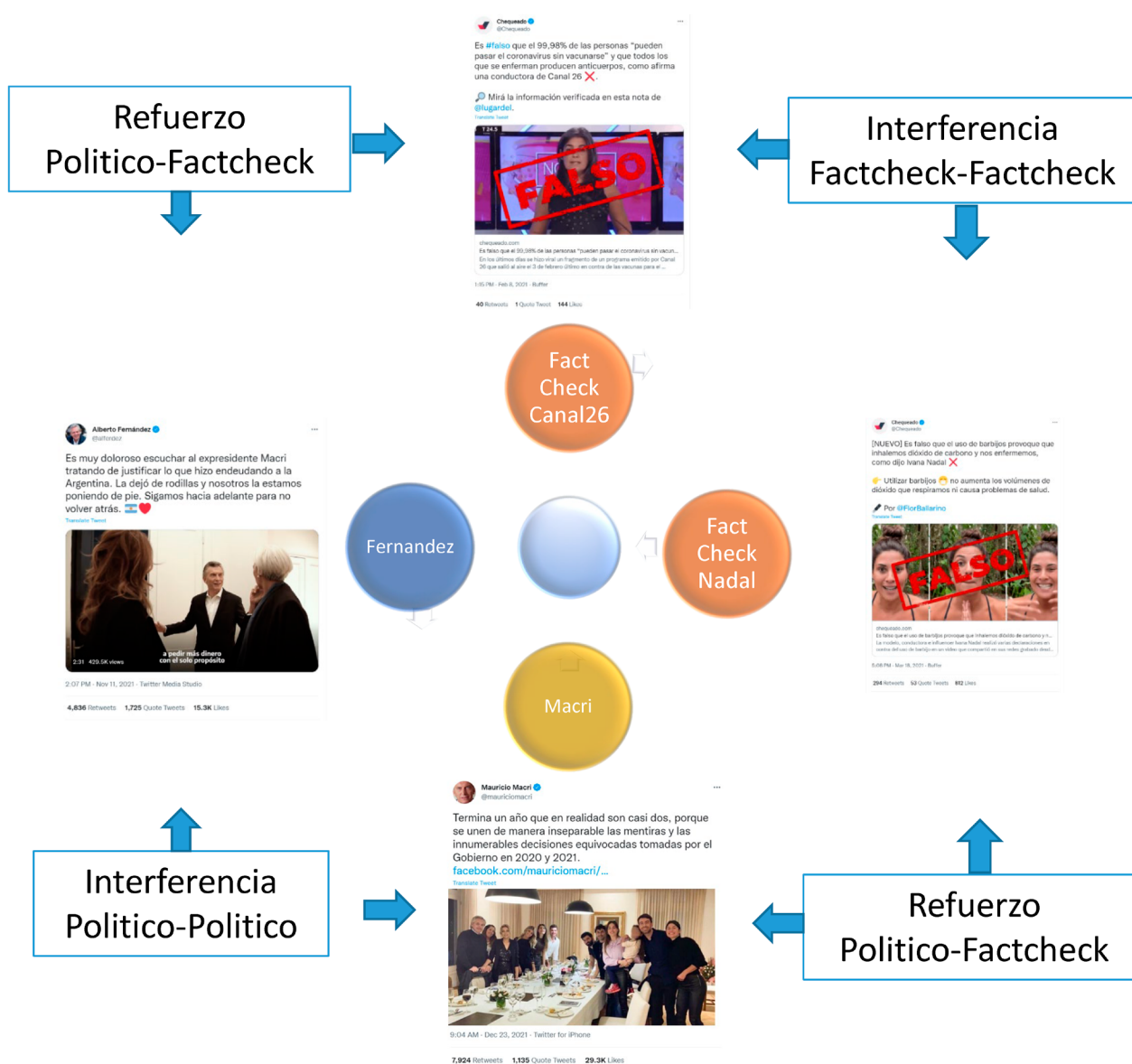
ANALYSIS

From Theory to Design

The primary aim of this study is to ascertain the nature of user responses to polarizing messages when examined within a contextual framework. Specifically, our focus involves evaluating the impact of fact-checking content when introduced into a broader context—presented alongside various messages, some of which carry distinct political polarizations. This inquiry aligns with a more comprehensive investigation into how diverse forms of political and non-political information influence the inclination to "like," share, and/or comment on content, contingent upon the engagement or disaffection engendered by the message.

The illustrative examples provided in Figure 2 offer insights into the executed experiment. The Chequeado tweet on the left, disseminated in March 2021, serves to debunk information propagated by influencer Ivana Nadal, asserting that masks used to prevent COVID-19 could elevate inhaled carbon dioxide levels, potentially leading to health issues. Various fact-checkers across Latin America discredited this claim and other anti-mask narratives circulating on social media. The tweet on the right, authored by Alberto Fernández in November 2021 amid Argentina's mid-term elections, involves Fernández criticizing

Figure 1.
Boosting and interference of paired messages



Note: Four different tweets. In our experiment, respondents are divided into six groups that alternately view Macri-Fernández, Macri-Chequeado1, Macri-Chequeado2, Fernández-Chequeado1, Fernández-Chequeado2, Chequeado1-Chequeado2. The probability of sharing Macri's tweet, Fernández's tweet, or one of the two Chequeado tweets changes depending on the combination presented to the respondents.

ing former President Mauricio Macri for escalating indebtedness levels during his tenure.

The second row of Figure 2 pairs the identical tweet from Alberto Fernández with another tweet from Mauricio Macri. In this combination, Macri criticizes Alberto Fernández recalling a dinner event where the latter breached quarantine protocols and provided misleading information—an image that gained widespread circulation in Argentina. After exposure to these paired tweets, respondents answer the following questions: (1) Which of these tweets are you more likely to SHARE on your wall? (Responses: tweet 1, tweet 2, both, none); (2) Which of these tweets are you

more likely NOT to SHARE on your wall? (Responses: tweet 1, tweet 2); (3) Which of these tweets do you expect to be commented on in your favorite news outlet? (Responses: tweet 1, tweet 2).

RESULTS

In digital platforms, our feed undergoes a continuous influx of novel posts, subjecting us to sequences of messages that are occasionally interrelated within a conversation and, at other times, swiftly succeeding one another, thereby modifying our interpretation of information. In social media networks, a fundamental rule prevails: messages presented in tandem are

Figure 2:
Alberto Fernández criticizes Mauricio Macri in its context



Tuit 1



Tuit 2



Tuit 1



Tuit 2

Note: In the different experimental treatments, Alberto Fernández's tweet was presented alternately alongside a Chequeado tweet or alongside a tweet from Mauricio Macri. We then asked respondents if they would share tweet 1, tweet 2, both, or neither.

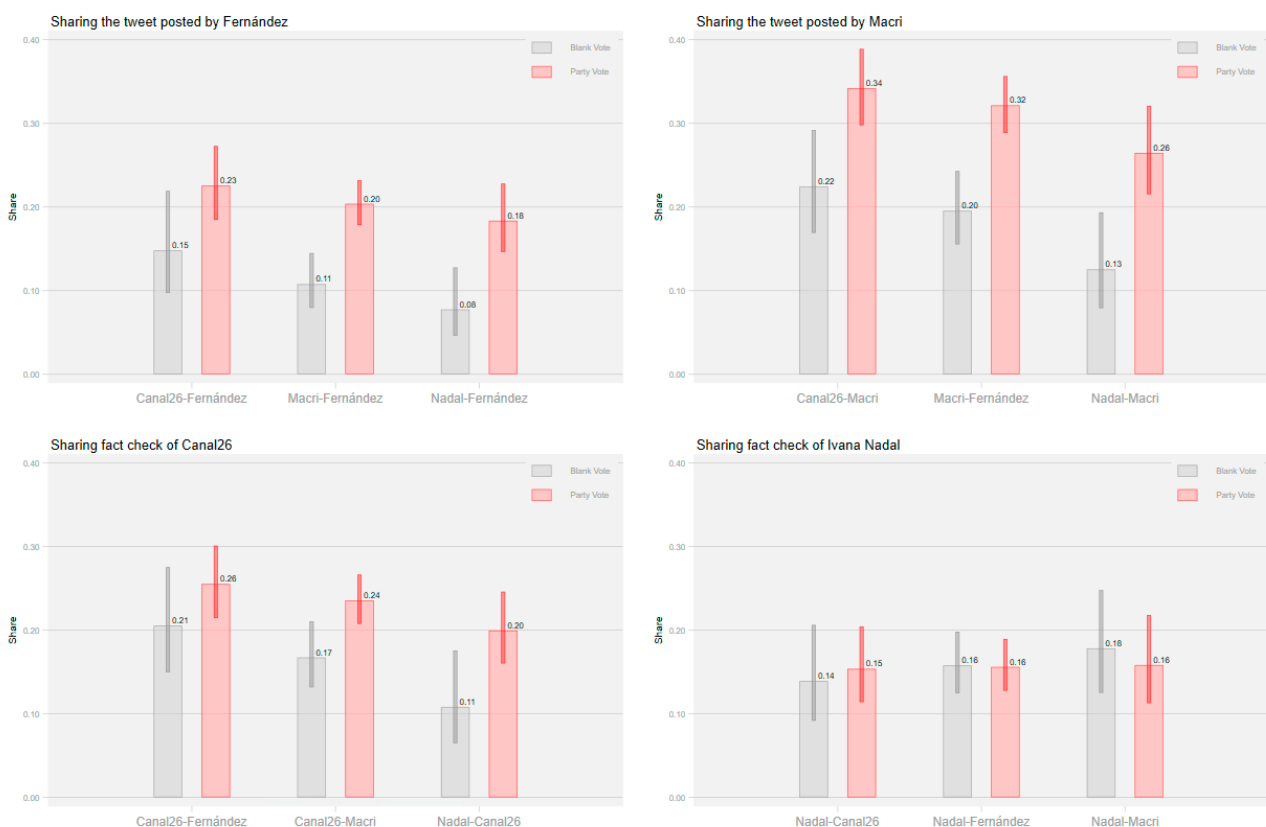
construed within the contextual framework, engaging in interactive dynamics and yielding reinterpretations of the primary content. This intricate interplay poses a formidable challenge for fact-checking endeavors in polarized environments, as the likelihood of a message being shared is contingent not merely upon how the verification is articulated—whether it is labeled as “false” or “true”—but also on the nature of the messages concurrently surfacing in users’ feeds.

Let us discuss why we expect variation in shared tweet rates when context changes. For instance, consider a scenario where a supporter of “Juntos por el Cambio” encounters a tweet from former President Mauricio Macri and simultaneously encounters a tweet from the fact-checker Chequeado, debunking misinformation. The click-through rate (CTR) of the “JxC” supporter is poised to ascend owing to the enthusiasm kindled by the perusal of Macri’s tweet, given his status as an authoritative figure within their in-group. We refer to the click-through rate as the number of times an advertisement is accessed from a platform (clicks) to the number of times it is dis-

played (impressions). Different companies purchase advertising space on Facebook, Twitter, or Instagram because each advertisement on a feed has a certain probability of a user clicking on it and, consequently, being transported to the seller’s website. The supplementary energy instigated by exposure to a pro-attitudinal message further extends its impact to all messages concomitantly displayed with Macri’s tweet, amplifying the probability of sharing both the political tweet and the correction.

Hypothesis 1 expected partisans to have a higher propensity to share tweets than non-partisans. This was expected for congruent messages, such as a Peronist sharing content for Alberto Fernández, and for fact-checks placed next to congruent tweets, such as sharing the correction in the top left of Figure 2. Consider Figure 3 below, which shows the rate at which messages are shared by partisans and non-partisans for each “primary” message as we rotate the other three tweets. In each plot, the red bars show the share rate by partisans, and the gray bars show the share rate by non-partisans. In 8 of the 12 pairings of tweets, partisans had a larger probability of sharing a

Figure 3
Sharing Twitter post by Partisans and Non-Partisans



Note: The sharing rate for respondents who voted for a party (partisans) or cast a blanc vote (non-partisans) is conditional on different combinations of political and fact-checking tweets. When presented together, the propensity to share a tweet increases for political and fact-checking tweets (constructive interference). In contrast, the intention to share decreases (destructive interference) when two fact-checking tweets are placed next to each other.

tweet than non-partisans at a statistically significant $p < .05$ level. In the other 4 cases, the difference in the sharing rate was not statistically significant. None of the 12 pairings had a higher propensity to share tweets by non-partisans.

Hypothesis 2 expected in-group partisans to share the tweet of their candidate at higher rates and, as important, to be activated to share non-partisan content. Figure 4 describes separate estimates for supporters of Alberto Fernandez (Blue), supporters of Mauricio Macri (Yellow), and blanc voters (gray). As expected, the most shared tweet by voters of Alberto Fernandez and Mauricio Macri was that posted by their candidates. More interesting, however, sharing of the fact-checks is higher for all voters when fact-checking tweets are paired with political tweets. In contrast, lower sharing takes place when fact-checking tweets are placed next to each other.

Within this section, we proffer two primary clusters of findings (See Figure 3). On the one hand, we explicate respondents' responses to tweets from the former President of "Frente de Todos," Alberto Fernández, and the former President of "Juntos por el Cambio," Mauricio Macri, when conjoined with other

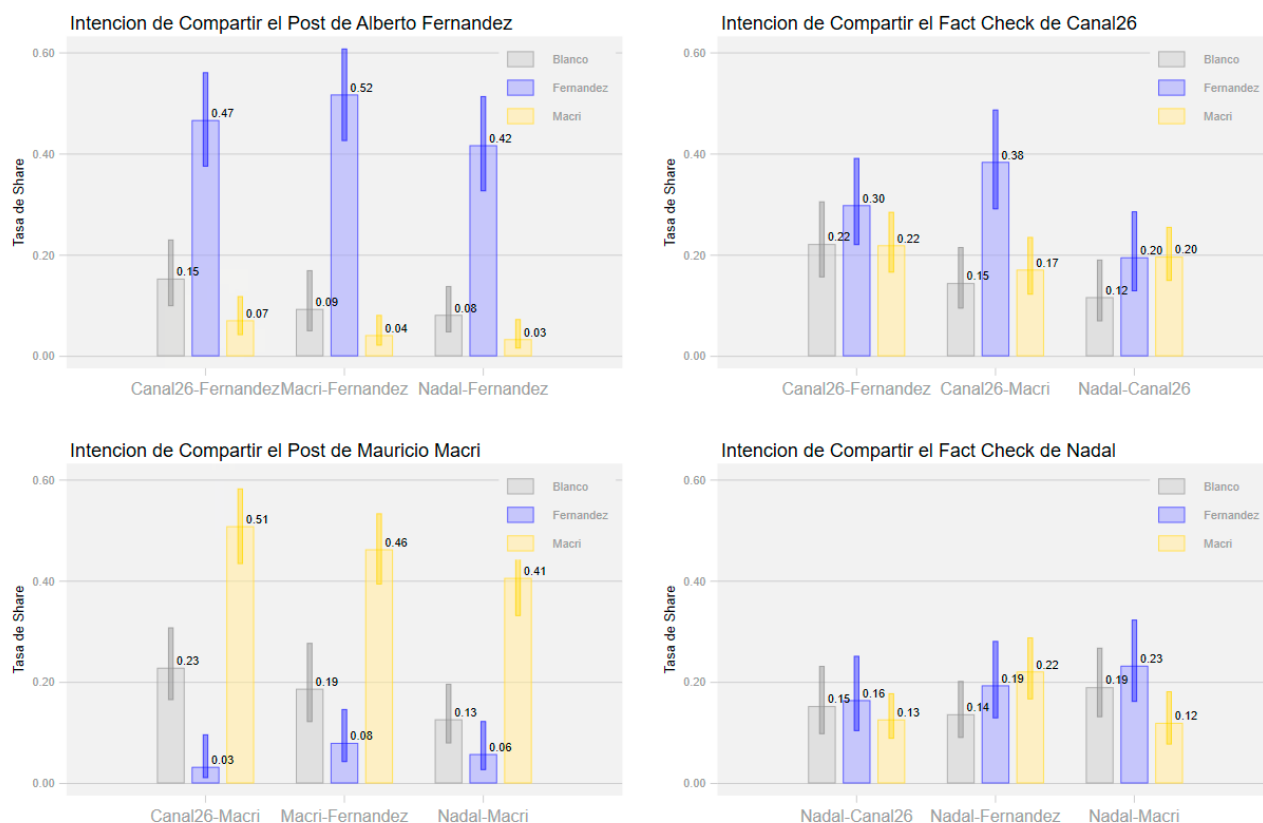
messages. These combinations are classified into **political-political** and **political-fact-check** categories, delineated in the left column. On the other hand, we delve into users' reactions to the two tweets from Chequeado when entwined with polarizing posts, distinguishing between **fact-check-fact-check** and **fact-check-political** combinations, elucidated in the right column.

Generally, when posts with political content are paired with fact-checks, there is reinforcement, meaning the click-through rate increases. Conversely, when two political tweets or two fact-checks are paired, interference decreases the intention to share.

The left graphs in Figure 4 show the sharing rate of Alberto Fernández's tweet when paired alternatively with Macri's post and the two Chequeado publications (Upper-left graph). Simultaneously, it displays the propensity to share Macri's post paired with Fernández's and the two fact-checks (Lower-left graph).

The first two hypotheses of this study posit that message activation is higher among partisan users, and this propensity to share through retweets increases with pro-attitudinal content. These assumptions are

Figure 4
Interactions in the Context of Political Tweets and Fact-Checking



Note: Sharing rate when respondents are exposed to paired political tweets and fact-checks. The propensity to share a tweet increases for political and fact-checking tweets when presented together (constructive interference). In contrast, the intention to share decreases (destructive interference) when two fact-checking tweets are placed next to each other.

confirmed in the two left graphs. Fernández's voters predominantly share the post from the former President of "Frente de Todos," followed by the Canal26 and Nadal fact-checks and to a lesser extent, Mauricio Macri's post. Similarly, users identified with Macrismo predominantly retweet Mauricio Macri, followed by Canal26 and Nadal, and lastly, Fernández. Ivanna Nadal's fact-check is the least shared of all.

Beyond confirming the widespread enthusiasm for sharing pro-attitudinal content among partisan peers, the activation of Macri's tweet by "Frente de Todos" respondents is relatively low. The additional negative energy of seeing an attack by Macri on Fernández increases their propensity to share the "other" alternative. Consequently, Canal 26's fact-check is retweeted more by Peronist voters when paired with Macri's tweet ("canal26-Macri") than when paired with Fernández's ("canal26-Fernández"). The sharing rate increases when Macri's tweet is paired with one of the fact-checks, while it decreases when respondents observe the political-political pairing.

In the graphs located on the right of Figure 4, Chequeado's posts are more shared when paired with a political tweet. Conversely, when the two fact-checks are shown together, the level of positive and negative energy drops, so they are shared less frequently. In other words, two Chequeado tweets interfere with each other, whereas pairing a political and a non-political tweet reinforces the sharing rate of both messages (See Figure 3).

We consider "reinforcement" to occur if the sharing rate increases when two messages are presented together. We believe "interference" is created if the sharing rate decreases when two of these possible messages are presented together. These findings contradict the last hypothesis of this study, which models the contextual effect of posts among the most partisan. Initially, we assumed that politically critical messages would increase activation among in-group supporters and that the sharing rate of fact-checks would also increase when they are together. Ultimately, the sharing rate of Chequeado's posts increases when "Frente de Todos" voters are negatively energized (feel disgust or anger for perceiving themselves as attacked), and it also grows when respondents identified with Macri are positively energized (with feelings of optimism and joy). For example, when the tweets of Ivana Nadal and Mauricio Macri are paired, "Frente de Todos" supporters share Nadal's correction more than when the pairing is between Nadal and Canal 26's fact-checks.

DISCUSSION

Results in this article show that the likelihood that users will share a particular post is affected by which content is presented next to it. Messages are never read in isolation but rather appear in a wall of other content that competes for attention and affects how we interpret the text, images, and authorship of social

media contents. Our experiment shows that, as we rotate contextual information, we change the decision to share content.

The effect of context on sharing is also sensitive to the users' partisan and non-partisan traits. Differences in content sharing that are driven by context allow us to clarify partisan effects in highly polarized social media environments and to measure the incidence of fact-checking interventions. The implementation of survey experiments with paired and rotating messages shows that disambiguation, which narrows the interpretation of a primary social media message, shapes how we interact with content and our decision to share. The outcomes of this investigation, therefore, validates recent scholarly inquiries on the topological effect of networks in confirmation bias (Aruquete & Calvo, 2023; Sikder et al., 2020).

Support for our first two hypotheses shows that party voters are activated by partisan messages, increasing the rate of sharing of social media posts that are pro-attitudinal but also the rate of sharing social media posts more generally, including fact-checks. Partisans have a greater propensity to disseminate content on social media, including political messages they like but also social media messages in general. This expectation was confirmed for both the incumbent party and the opposition, and aligns well with recent work on the incidence of partisan attitudes on the formation of bubbles (Aruquete & Calvo, 2023; Sikder et al., 2020).

Support for the last two hypotheses of this study is more novel and interesting: we show that sharing a primary post is affected by the presence of different secondary posts—context matters, both for sharing posts and for the particular case of fact-checks. As expected, polarizing political messages increase the propensity to share in-group messages, but also fact-checks that are shown next to the partisan messages. Our experiment contributes to the fact-checking literature by showing that two fact-checks shown together are less likely to be shared—destructive interference—while pairing political and non-political messages increases sharing-constructive interference.

LIMITATIONS OF THE STUDY

Every study has its limitations, and it is crucial to highlight the blind spots in an experimental analysis of network content activation (NAF). This section emphasizes two important methodological considerations: external validity and network selection.

The decision to experimentally evaluate the propensity to share content on social media—whether the messages are true, false, or corrections—offers many comparative advantages over studies using observational data. The most significant advantage is the internal validity resulting from experimental con-

trol over what a respondent observes. However, the higher internal validity of experiments can be accompanied by lower external validity, as the experimental information is presented to users in a context different from their everyday experiences. On Facebook, for instance, posts appear sequentially (top/bottom) rather than in competition (left/right). More importantly, presenting two tweets isolated from the context of a feed increases user attention. Consequently, the rate of “likes” and “retweets” in experimental studies is higher than in observational data.

Furthermore, in its experimental version, this type of study appropriately evaluates normative congruence but exaggerates the activation resulting from that congruence. Although it has achieved significant results in this field of study, our experiment presents two limitations regarding the study’s external validity. The first limitation is that while the experiment has internal validity, it lacks external validity for two reasons. First, the exposure to and attention given to the content is greater in survey situations than in the usual social media context, where different messages compete for users’ attention. Therefore, the share rate observed in this experiment will also be higher than in the usual social media context. Second, in the experiment, the behavioral response is measured after the treatment—exposing users to the content—to capture the respondents’ reading and reaction times. Unlike the responses obtained after the treatment, on social media, the reaction occurs simultaneously with the exposure to the messages. Regarding the second limitation, the affective response of the respondents is self-reported. Although this is a standard tool for measuring responses in surveys, individuals may report “desirable” emotions that differ from the affective response they experience under normal conditions when interacting on social media.

CONFLICT OF INTEREST STATEMENT

The authors of this article declare that they have no financial, professional or personal

conflicts of interest that could have inappropriately influenced this work.

AUTHORSHIP CONTRIBUTION STATEMENT

Natalia Aruguete: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review.

Ernesto Calvo: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review.

REFERENCES

- Aruguete, Natalia and Calvo, Ernesto. 2023. *Nosotros contra ellos. Cómo trabajan las redes para confirmar nuestras creencias y rechazar las de los otros*. Buenos Aires: Siglo XXI.
- Aruguete, Natalia. 2021. “Activación de encuadres en red. Un modelo para repensar la circulación de sentidos en el nuevo entorno mediático.” *Profesional De La información* 30(2): 1-18. <https://doi.org/10.3145/epi.2021.mar.18>
- Aruguete, Natalia and Calvo, Ernesto. 2018. “Time to #protest: Selective exposure, cascading activation, and framing in social media.” *Journal of Communication*, 68(3): 480-502. <https://doi.org/10.1093/joc/jqy007>
- Aruguete, Natalia, Calvo, Ernesto, and Ventura, Tiago. 2023. “Network activated frames: content sharing and perceived polarization in social media.” *Journal of Communication*, 73(1): 14-24. <https://doi.org/10.1093/joc/jqac035>
- Barberá, Pablo et al. 2015. “The critical periphery in the growth of social protests.” *PloS One*, 10(11): 1-15. e0143611. <https://doi.org/10.1371/journal.pone.0143611>
- Bélanger, Éric, and Meguid, Bonnie M. 2008. “Issue salience, issue ownership, and issue-based vote choice.” *Electoral Studies*, 27(3): 477-491. <https://doi.org/10.1016/j.electstud.2008.01.001>
- Boczkowski, Pablo J., Mitchelstein, Eugenia, and Matassi, Mora. 2018. “News comes across when I’m in a moment of leisure: Understanding the practices of incidental news consumption on social media.” *New media & Society*, 20(10):3523-3539. <https://doi.org/10.1177/1461444817750396>
- Cruz, Facundo. 2021. “Cuando la grieta baja desde arriba. Bicoalicionismo y competencia política polarizada en Argentina.” In Ramírez, Ignacio and Quevedo, Alberto. *Por qué preferimos la grieta (aunque digamos lo contrario)*. (pp. 103-134) Buenos Aires: Capital Intelectual.
- Druckman, James N., et al. 2021. “Affective polarization, local contexts and public opinion in America.” *Nature Human Behavior*, 5(1): 28-38.
- Epstein, Lee, and Segal, Jeffrey A. 2000. “Measuring issue salience.” *American Journal of Political Science*: 66-83. <https://doi.org/10.2307/2669293>
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven-London: Yale University Press.
- Franklin, Charles H., and Jackson, John E.. 1983. “The dynamics of party identification.” *American Political Science Review*, 77 (4): 957-973.
- Gilbert, Daniel. 2006. *Tropezar con la felicidad*. Barcelona: Destination.
- Hainmueller, Jens, Hopkins, Daniel J., and Yamamoto, Teppei. 2014. “Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments.” *Political analysis*, 22(1): 1-30. <https://doi.org/10.1093/pan/mpt024>
- Huddy, Leonie, Mason, Liliana, and Aarøe, Lene. 2015. “Expressive partisanship: Campaign involvement, political emotion, and partisan identity.” *American Political Science Review*, 109 (1): 1-17. <https://doi.org/10.1017/S0003055414000604>
- Iyengar, Shanto, & Westwood, Sean J. 2015. “Fear and loathing across party lines: New evidence on group polarization.” *American Journal of Political Science*, 59 (3): 690-707. <https://doi.org/10.1111/ajps.12152>
- Iyengar, Shanto et al. 2019. “The origins and consequences of affective polarization in the United States.” *Annual review of political science*, 22: 129-146. <https://doi.org/10.1146/annurev-polisci-051117-073034>
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.
- Kaplan, Noah, Park, David K., and Ridout, Travis N. 2006. “Dialogue in american political campaigns? An examination of issue convergence in candidate television advertising.”

- American Journal of Political Science*, 50 (3): 724–736. <https://doi.org/10.1111/j.1540-5907.2006.00212.x>
- Leeper, Thomas J., Hobolt, Sara B., and Tilley, James. 2020. "Measuring subgroup preferences in conjoint experiments." *Political Analysis*, 28(2): 207-221. <https://doi.org/10.1017/pan.2019.30>
- Mason, Liliana. 2015. "I disrespectfully agree: The differential effects of partisan sorting on social and issue polarization." *American Journal of Political Science*, 59(1): 128-145. <https://doi.org/10.1111/ajps.12089>
- Mason, Liliana. 2016. "A cross-cutting calm: How social sorting drives affective polarization." *Public Opinion Quarterly*, 80 (S1), 351-377. <https://doi.org/10.1093/poq/nfw001>
- McCoy Mason, Jennifer, Rahman, Tahmina, and Somer, Murat. 2018. Polarization and the global crisis of democracy: Common patterns, dynamics, and pernicious consequences for democratic politics." *American Behavioral Scientist*, 62(1): 16-42. <https://doi.org/10.1177/0002764218759576>
- Petrocik, John R. 1996. "Issue ownership in presidential elections, with a 1980 case study." *American Journal of Political Science*, 40(3): 825-850.
- Organización Panamericana de la Salud. 2020. *Entender la infodemia y la desinformación en la lucha contra la COVID-19*. Recuperado de: <https://iris.paho.org/handle/10665.2/52053>.
- Saxena, Bhawna, & Kumar, Padam. 2019. "A node activity and connectivity-based model for influence maximization in social networks." *Social Network Analysis and Mining*, 9: 1-16. <https://doi.org/10.1007/s13278-019-0586-6>
- Schulíaquer, Ivan and Vommaro, Gabriel. 2020. "Introducción: La polarización política, los medios y las redes. Coordinadas de una agenda en construcción." *Revista Saap*, 14(2): 235-247.
- Sikder, Orowa et al. 2020. "A minimalistic model of bias, polarization and misinformation in social networks." *Scientific reports*, 10(1): 5493. <https://doi.org/10.1038/s41598-020-62085-w>
- Webster, Steven W., and Abramowitz, Alan I. 2017. "The ideological foundations of affective polarization in the US electorate." *American Politics Research*, 45 (4): 621-647. <https://doi.org/10.1177/1532673X17703132>