

Do Brands Matter? Understanding Public Trust in Third-Party Factcheckers of Misinformation and Disinformation on Facebook

ANDREA CARSON*¹

La Trobe University, Australia

TIMOTHY B. GRAVELLE

Bain and Company, Canada
Wilfrid Laurier University, Canada

JUSTIN B. PHILLIPS

University of Waikato, New Zealand

JAMES MEESE

RMIT University, Australia

LEAH RUPPANNER

University of Melbourne, Australia

The spread of misinformation and disinformation is an urgent global problem threatening information quality. Third-party fact checking is widely used to mitigate its harmful effects. Yet, the relationship between fact checking and misinformation spread is understudied. This study addresses this gap and investigates public trust in fact checkers and engagement with debunked claims. Drawing on the theory of motivated reasoning, we use real-life

Andrea Carson: a.carson@latrobe.edu.au

Timothy B. Gravelle: tgravelle@wlu.ca

Justin B. Phillips: justin.phillips@waikato.ac.nz

James Meese: james.meese@rmit.edu.au

Leah Ruppanner: leah.ruppanner@unimelb.edu.au

Date submitted: 2023-01-27

¹ We would like to acknowledge this research was funded by a global Meta grant to study mis- and disinformation. Meta played no role in any aspect of the research design or study findings. This research has human subjects ethics approval from La Trobe University, Melbourne, Australia. Ethics approval number: HEC21393. We are grateful to the reviewers for their very helpful feedback to improve this paper. We also thank Austin Pettis and Jon Cohen at SurveyMonkey for their support in building the SurveyMonkey endpage technology for Australia, making it possible to run our survey-based experiment. Replication code, data, survey questionnaire, and appendix materials for this article are available online in an Open Science Framework archive, available at: <https://doi.org/10.17605/OSF.IO/SN59H>.

Copyright © 2023 (Andrea Carson, Timothy B. Gravelle, Justin B. Phillips, James Meese, and Leah Ruppanner). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

disinformation about former Prime Minister Scott Morrison's response to flood victims during the 2022 Australian election. We undertake a survey experiment ($N = 8,235$) and alter the fact-check source to measure public trust and subsequent engagement with disinformation. Overall, we find high trust in fact checking. However, we also find a third of participants will likely engage with disinformation despite trusting a fact check that explicitly states it is false. Our study lends support to motivated reasoning, finding a disconnect between trust in fact checkers and their capacity to limit disinformation spread.

Keywords: misinformation, disinformation, third-party fact checking, motivated reasoning, Facebook

The online spread of "fake news," such as misinformation and disinformation, is a significant global problem that can erode public trust and harm core pillars of modern society. Misinformation is defined here as the spread of inaccurate or misleading content whether harmful or not (Gibbons & Carson, 2022). Disinformation, on the other hand, is defined as the spread of misleading, inaccurate, or deceptive content through decisive actions that lead to harm (Gibbons & Carson, 2022). Both misinformation and disinformation can cause real-world harm, if left unchecked. Together, misinformation and disinformation fall under the umbrella term of "fake news," and although this phenomenon is not new, the digital affordances of the Internet age make its rapid, vast spread unprecedented. Social cohesion, public health and safety, political stability, electoral integrity, and, ultimately, democracy are under threat by its local and global spread. The COVID-19 pandemic has further illustrated how misinformation and disinformation about causes and treatments of the novel coronavirus can have deadly consequences (Andrews & Paquette, 2020; Shepherd, 2020).

Considering the speed of dissemination and potential for real-world harm, limiting the spread and harms of misinformation and disinformation has become a first-order concern for governments, policymakers, and stakeholders (Meese & Hurcombe, 2020). Studies internationally, including from our case country of Australia, show citizens are worried about fake news and their own capacities to distinguish fact from fiction (Barthel, Mitchell, & Holcomb, 2016; Park, Fischer, Flew, & Dulleck, 2020). Leading scholars have forewarned of an epistemic crisis where people become unable to decipher fact from fiction (see Dahlgren, 2018). Mitigation measures have been deployed to deal with this urgent global problem, including preemptive (prebunking) and reactive (debunking) correction programs (Carson, Gibbons, Martin, & Phillips, 2022). The focus of this study is on debunking false claims using third-party fact checkers. Our aim is to assess public trust in third-party fact checkers to arbitrate contentious claims and their capacity to limit the spread of online misinformation and disinformation.

Like fake news, fact checking is not a new phenomenon. Traditional media outlets have historically undertaken their own internal fact checks. What is novel this century is the emergence of third-party fact checking to mitigate the widespread proliferation of misinformation and disinformation (Amazeen, Thorson, Muddiman, & Graves, 2018; Graves, 2016). We use the term third party to refer to fact checking of contentious claims by an organization that is separate to the outlet that first published the claims. Third-party fact checking emerged in 2003 with FactCheck.org, a nonprofit, nonpartisan U.S. university that focused on verifying political claims. Since then, third-party fact checking has expanded to investigate public

claims beyond politics including contentious claims in consumer news, health, and advertising. Fischer (2020) reports more than 300 fact checkers are operating worldwide to form a multibillion-dollar industry.

As third-party fact checkers have proliferated, so too have the number of fact-check studies. Much of this scholarship has focused on the effectiveness of fact checking, both in its efficiency in correcting false information and the impact of fact checking on changing public perceptions (Nieminen & Rapeli, 2019; Owen et al., 2019; Walter, Cohen, Holbert, & Morag, 2020). This persistent interest in whether fact checking works is important given that policymakers, journalists, and digital platforms generally view the intervention as useful and worthy of investment. Yet, we know little about whether public trust in different fact-checker brands differs, or even if it matters in limiting the spread of debunked online falsehoods.

This study directly fills this void using an experimental survey design to investigate public trust in leading fact-checking brands in Australia and Australians' subsequent engagement with a debunked political claim. Drawing on the theory of motivated reasoning in a political context (discussed below) and literature on source credibility, we examine whether the source of a fact-check structures public trust in fact checking and shapes how people respond to a correction. Our aim is to understand if people consider fact-checking brands equally by testing how people respond to the same real-life fact check that has been rebranded for each treatment group. Further, we investigate public engagement with false information after it is corrected. This is important to understand because social media engagement on Facebook, through likes, dislikes, shares, and so forth, can stimulate the algorithmic spread of content, thus undermining fact-checkers' capacity to limit its spread.

This article proceeds as follows. We outline the extant fact-checking scholarship before discussing the theory of motivated reasoning and related research. We draw on source credibility literature and provide context for Australian-specific fact checkers. We then explain our study design and its findings to conclude with a discussion of their implications and recommendations.

Literature Review: The Efficacy and Limitations of Fact Checking

Existing studies investigate fact-checking efficacy in varied contexts. An extensive review of fact-checking research presents "mixed results" on its efficacy to disabuse people of misinformation and from sharing it (Nieminen & Rapeli, 2019, p. 297).

Of the different research streams, one compares the effectiveness of fact-checking formats (such as the use of ratings, scales, "false/true" tags) or their tone (humorous vs. nonhumorous; Amazeen et al., 2018; Garrett, Nisbet, & Lynch, 2013) to correct online falsehoods. Fact checks presented to research participants often come from one particular fact-checking institution (Garrett et al., 2013) and use nonpartisan sources (Jarman, 2016). Another stream has focused on the role of fact checking in political contexts, such as whether summary fact checking changes audiences' perceptions of politicians (Agadjanian et al., 2019) and election campaign debates online (Coddington, Molyneux, & Lawrence, 2014). In all of these studies, the focus is on the *outcome* of the fact check (e.g., whether the fact check is effective and under what conditions), rather than the *input*—how the public views the fact checker.

In such studies, fact-checking organizations are often treated as objective and neutral actors against which various aspects of a corrective intervention can be tested (Graves, 2017). A prominent Australian third-party fact checker, Australian Associated Press (AAP, 2023) supports this proposition and explicitly states its commitment to political neutrality on its website (AAP, 2023). The common assumption that fact checkers are neutral may explain the paucity of research comparing public perceptions of trust between fact checkers. Given fact-checking's antecedents in correcting political statements (i.e., FactCheck.org; PolitiFact) it is reasonable to question if the public view fact checkers through a partisan lens or exhibit bias toward the fact-checking source. To test this, we use a real-world-like scenario to understand differences in the input, in this case trust between Australian third-party fact-checking brands and subsequent public engagement with debunked disinformation.

A tacit assumption in fact checking is that false beliefs can be corrected by just providing relevant facts, known in psychology as the information deficit model (IDM). But as has been documented, this model, reliant on rational calculations, has shortcomings and does not adequately explain other cognitive, social, and affective drivers of people's attitude formation and truth judgments (Ecker et al., 2022). For example, a person can have the "facts" and be aware of scientific consensus on vaccines but still hold antivaccination views (see Simis, Madden, Cacciato, & Yeo, 2016). Political or ideological orientations can be one factor influencing belief formation with scholars studying the role that political leanings can play when it comes to people accepting or rejecting fact-check corrections (Jarman, 2016; Shin & Thorson, 2017; Walter et al., 2020). Some find correcting misinformation with straight facts can have a limited effect on people with strong ideological worldviews (Nyhan & Reifler, 2010; Nyhan, Reifler, & Ubel, 2013). Nyhan and Reifler (2010) found that "corrections frequently fail to reduce misperceptions among the targeted ideological group" (p. 303).

Motivated Reasoning and Backfire Effects

Taber and Lodge (2006) examined motivated reasoning by testing what citizens were *motivated* to do with information rather than simply their levels of information acquisition (i.e., IDM). They found people often did not integrate new information in an unbiased way. Since then, researchers have applied this work to fact checking and some have identified a "backfire effect," whereby "corrections actually *increase* [*sic*] misperceptions among the group in question" (Nyhan & Reifler, 2010, p. 303). Attributed to the theory of motivated reasoning, such studies find that people are motivated to disbelieve or ignore certain types of information that contradict existing ideological or partisan beliefs. Scholars have found that motivated reasoning is more likely during election campaigns when political attention intensifies leading to resistance to fact checks that contradict preexisting political views (Lenz, 2012; Walter et al., 2020).

Relatedly is "confirmation bias," which refers to how people favor information that is consistent with their existing beliefs (Lazer et al., 2018). Thus, people tend to trust a message when it aligns to their preexisting political views, notwithstanding its accuracy. These studies suggest that fact-checking political disinformation may be ineffective, especially among those who hold strong ideological commitments. However, as indicated earlier, fact-checking research has produced mixed findings, including on backfire effects. Studies examining different fact-check formats (Ecker, O'Reilly, Reid, & Chang, 2020) and active journalistic adjudication of information (avoiding "he said/she said" journalism; Pingree, Hill, & McLeod,

2013) find little support for backfire effects, as did a meta-analysis of fact-checking studies (Swire-Thompson, DeGutis, & Lazer, 2020). Considering these mixed findings, our study is exploratory, drawing on this literature to inform our first two of three research questions:

RQ1: Does fact checking reduce trust in an incorrect Facebook post?

RQ2: What factors (fact-check source; individual political orientations) influence trust in a fact check?

Public Engagement With Online Falsehoods

We also aim to understand how people engage with debunked information once it is fact checked. As discussed, in some circumstances, misinformation can continue to influence people's thinking and engagement even after they receive a correction, and even after they accept that the correction is true (Carson et al., 2022; Nyhan & Reifler, 2010). For example, a Pew Research Center survey found 14% of U.S. respondents admitted to knowingly sharing false news (Barthel et al., 2016). This persistence is called the continued influence effect (CIE). Multifold drivers of CIE have been identified including conspiratorial mentality, fear, identity expression, and motivated reasoning (Ecker et al., 2022).

Identity expression, which can include political identification, has also been found to increase engagement and sharing of misinformation to signal group membership (Brady, Gantman, & Van Bavel, 2020). Studies find people are willing to trade off accuracy to build social connections when making content sharing decisions and, in some cases, expect the sharing of falsehoods to generate higher social value than sharing factual news (Ren, Dimant, & Schweitzer, 2023). Others have been shown to share misinformation to fuel moral outrage in others (Brady et al., 2020).

Scholars who have explored CIE factors have found time spent on social media is positively linked to how much that individual consumes fake news (Nelson & Taneja, 2018). For some individuals, repeated exposure to online falsehoods can function as a heuristic, leading them to believe that information, seen repeatedly, might be true (Pennycook, Cannon, & Rand, 2018).

Additionally, engaging with and sharing information on social media is often dictated by what captures one's attention. Research finds that moral-emotional words such as "fight," "greed," "evil," and "punish" are prioritized in early visual attention over other arousing words and can lead to increased sharing (Ecker et al., 2022; see also Brady et al., 2020). In political communication, adding a single moral-emotional word to tweets about contentious political issues such as gun control has been found to increase retweets by 20% (Brady, Wills, Jost, Tucker, & Van Bavel, 2017). Thus, emotion can play a role in the tendency to believe and/or share misinformation and disinformation. With these findings in mind, we use a real-world emotive example of disinformation (the accusation that Queensland flood victims are weaponizing their trauma against the Australian prime minister) in a politically charged setting (a federal election campaign), to understand the propensity to share emotive political disinformation after it has been debunked. This leads to our final research question:

RQ3: Does a third-party fact check debunking a false political claim reduce the intention of social media users to react to that false claim on Facebook?

Source Credibility

Source credibility provides important social cues that influence belief formation. Third-party fact checking usually undertaken by professional journalists offers a form of source credibility to help navigate misinformation and disinformation. In general, Ecker and colleagues argue that “messages are more persuasive and seem to be true when they come from sources perceived to be credible rather than non-credible” (Ecker et al., 2022, p. 14). However, other researchers find that there can be a partisan element to this with conservatives less likely to adopt ideologically inconsistent information (Robertson, Mourão, & Thorson, 2020; Walter et al., 2020). Robertson et al.’s (2020) analysis of survey data after the 2016 U.S. presidential election shows that liberals (Democrat supporters) and liberal/mainstream news consumers have more favorable attitudes toward third-party fact-checking sites and will more readily use them compared with conservatives (Republican supporters). They found “results point to U.S. fact-checking sites being absorbed into wider ideological discourses and patterns of ideological news consumption” (Robertson et al., 2020, p. 217).

No such studies have been undertaken in the partisan use of fact checkers in the Australian environment. However, we do know that in terms of public attitudes to the Australian media, conservative voters are less trusting of professional journalists than are progressive voters (Carson et al., 2022; Jackman & Ratcliff, 2018). Studies also single out the national broadcaster. Surveys show that although the ABC overall is highly trusted by the population, it is less trusted than commercial media by conservative voters (Jackman & Ratcliff, 2018). This raises an interesting question about how fact checking will be regarded by Australian conservatives when the source partially contains ABC branding as in RMIT ABC Factcheck.

Fact Checkers in Australia

Australia provides a powerful case to understand source credibility of fact checkers. It has three domestic fact-checking organizations in operation, Australian Associated Press (AAP), RMIT ABC Factcheck, and RMIT FactLab. All, at the time of the experiment, had some external credibility as signatories to the fact-checking accreditation body, the International Factchecking Network (IFCN), which requires members to adhere to prescribed standards. In addition, operating in the Australian fact-checking landscape is a small presence of international wire services such as Reuters Factcheck and France’s AFP Australia Factcheck.

Of the domestic outlets, AAP is a national newswire service established in 1935 that has developed a reputation as an “independent” news source by providing news content to competing newsrooms from across the political divide (Ellis, 2010) and added fact checking to its remit in recent years.

In the past decade, two additional fact-checking organizations have emerged, RMIT ABC Fact Check and RMIT FactLab, operating with a similar staffing base as AAP to undertake specific fact-checking roles. RMIT ABC Fact Check’s (2023) mission is to fact check “claims by public figures” (p. 1), whereas RMIT FactLab fact checks content specifically from social media and at the time of the experiment received funding from Meta to do so. RMIT FactLab is housed within a university (Royal Melbourne Institute of Technology). The careful separation of the Australian Broadcasting Corporation (ABC) from performing social media

corrections of Facebook gestures to the ideological and partisan challenges surrounding third-party fact checking identified above and examined in this study.

The ABC is a national public service media organization in Australia with a broad mission to provide “news, current affairs, entertainment and cultural enrichment” across broadcast and digital media services within and outside Australia (Australian Government, 1983, p. 7). Despite its commitment to public service journalism, right-wing critics regularly accuse it of favoring left-wing views and failing to promote conservative opinions (Miragliotta & Errington, 2012). Although RMIT ABC Fact Check’s charter is as an objective arbiter, as recognized by its IFCN accreditation, these political contestations show how difficult it can be for third-party fact checkers to be *seen* as objective and politically neutral. The extent to which right-wing audiences actually trust fact checks coming from an ABC-affiliated outlet is explored here.

This inquiry into public trust in fact-checks is critical given a growing scholarship documenting citizens are losing trust in related media organizations in democracies (Ardèvol-Abreu & Gil de Zúñiga, 2017; Fisher, 2019; Fletcher & Park, 2017; Robinson, 2018). In this study, we expand the focus beyond media trust to third-party fact checkers. Additionally, we present these fact checks within a platform that is sometimes used to share misinformation—Meta’s Facebook. As Park and colleagues (2020) note, a growing body of research is finding “an overall skepticism or suspicion around information that news audiences encounter on digital platforms” (p. 84; see also Fletcher & Nielsen, 2019). Here, we test whether fact checks from different third-party fact checkers on Meta’s Facebook platform influence trust in the information and, critically, the likelihood that respondents will share known false information. This allows us to contribute to scholarship on the effectiveness of third-party fact checking to disabuse citizens of misinformation and disinformation, and to understand Facebook users’ propensity to share known false content.

Data and Methods

To answer our three research questions, we exploit a real-life example of a TV screenshot that was altered to encompass a false, politically charged caption, which is then shared on social media, including Facebook. We deviate from past studies that use a single fact-check brand and conduct a survey experiment that assesses participants’ reaction to the real-life like correction (of the altered TV image posted on Facebook) undertaken by three prominent Australian fact-checking units: AAP, RMIT FactLab, and RMIT ABC Fact Check. We add a fourth international fact checker, Reuters Fact Check, because it has no obvious partisan history in Australia and it also fact checked the same disinformation post.

We evaluate our theoretical expectations about fact checking with an online survey containing an embedded experiment ($N = 8,235$). Survey data collection took place between May 6, 2022, and May 20, 2022, during the 2022 Australian federal election campaign. The survey was conducted by SurveyMonkey using their “endpage” methodology—a type of online river sample. This involves recruiting online survey respondents from survey-takers who complete one of the thousands of user-created surveys on SurveyMonkey’s platform every day. After completing a survey, a sample of respondents from Australia (determined from their Internet protocol [IP] addresses) were presented with a survey completion Web page (“endpage”) inviting them to complete another (voluntary, uncompensated) survey on current events.

This sampling methodology yields large, demographically representative samples that have been used to study a range of topics relating to public opinion, voting behavior, consumer behavior, and public health (Chen, Valliant, & Elliott, 2019; Gravelle, 2021; Gravelle, Phillips, Reifler, & Scotto, 2022; Williams, Gravelle, & Klar, 2022).

Our survey data constitute a nonprobability sample, though they are nevertheless well suited to our research aims (cf. Baker et al., 2013). Users of Facebook and other social media platforms are online by definition, mitigating noncoverage bias because of the exclusion of the offline population. Nevertheless, it is important to underline that our focus is on making inferences about the causal effects of our experimental manipulations and not descriptive inferences about the Australian population writ large (cf. Druckman, 2022; Mutz, 2011). Further, research comparing survey experiments conducted with online nonprobability samples and probability samples have found that they yield very similar estimates of treatment effects (Coppock, Leeper, & Mullinix, 2018; Krupnikov & Levine, 2014; Mullinix, Leeper, Druckman, & Freese, 2015).

The key elements of our survey design are as follows (the complete survey instrument is included in the replication archive). The survey began with introductory questions measuring respondents' opinions about whether Australia was heading in the right or wrong direction, which current issue mattered most, general interest in politics, and interest in the current federal election. The survey asked respondents to report which social media platforms they used and the frequency with which they used specific platforms. Respondents were then asked to rate the extent to which they liked (or disliked) the major Australian political parties and their respective leaders. Following these introductory questions, the subset of respondents who reported using Facebook were presented with a randomly assigned experimental condition.

All Facebook users, including those in the control condition ($n = 1,770$), were presented with an actual (false) Facebook post with an image from a television broadcast and a text description where then Prime Minister Scott Morrison purportedly criticized victims of flooding in Queensland for "weaponising their trauma" (see RMIT FactLab, 2022) against the incumbent Liberal-National Coalition government (see Figure 1).

BREAKING:
 Morrison complains that flood victims
 are "weaponising their trauma"
 against the government.
[#auspol](#) [#floods2022](#)



Figure 1. A real manipulated-image involving then Australian Prime Minister Scott Morrison, distributed on social media platforms including Facebook during the 2022 Australian federal election. Source: see RMIT FactLab (2022).

This image was altered (by an unknown party) by providing a fabricated caption that distorts the context. It was circulated on Facebook and other social media platforms in March 2022 before being fact checked and eventually removed. Unlike actual Facebook posts, our experiment did not attribute the image to an author. This intentionally departs from the user experience on the Facebook platform. Our aim here is to avoid confounding the effects of the post itself and the author of the post. Further, close adherence to the Facebook user experience is far less important in our study than the careful manipulation of our experimental treatments for causal inference (Druckman, 2022; Mutz, 2011).

Facebook users in our fact-check treatments were also exposed to a fact check attributed to one of four third-party fact checkers that had, in real-life, fact checked the exact image and pronounced the post as false (see Figure 2 for our standardized version using AAP). The text remained the same for all fact-check treatments. These were randomized and presented to the treatment groups: Australian Associated Press, AAP ($n = 1,581$), RMIT FactLab, ($n = 1,642$), RMIT ABC Fact Check ($n = 1,645$), and the international wire service, Reuters Fact Check ($n = 1,597$).



Figure 2. Sample of one of the four fact-checking treatments randomly assigned to participants. Source: Authors using abridged text from real-life fact checks of the “flood victim” claim.

Following this experimentally manipulated treatment, all Facebook-using respondents were asked the four questions that serve as our dependent variables. They were first asked the following question about the trustworthiness of the altered Facebook post: “How trustworthy is the Facebook post to you?” with the five-point response scale ranging from “Extremely trustworthy” to “Not at all trustworthy.” Next, those in the fact-check treatments were also asked about the trustworthiness of the fact check itself with the following question: “How trustworthy is the [organization] fact check to you?” with the same “Extremely trustworthy” to “Not at all trustworthy” response scale. All respondents were asked how they would respond to the Facebook post with the altered image—that is, how they would engage with the post on Facebook, whether by sharing it, commenting on it, clicking on one of Facebook emoji reactions (like, love, care, haha, wow, sad, angry), or doing none of these. From these response options, we create a dichotomous variable indicating any response to or engagement with Facebook (versus no response). Depending on whether respondents reported that they would react (or not) to the Facebook post, they were also presented with a follow-up, open-ended question where they were asked to describe in their own words why they would (or would not) respond to the Facebook post in that way.

Following the dependent variables, all respondents (including both Facebook users and nonusers) were asked about their voting intentions in the federal election held May 21, 2022, political party identification, self-rated ideological placement (from left to right), and their demographic characteristics, including sex, age, educational attainment, country of birth, state of residence, indigenous ancestry, and language most often spoken at home.

For ease of interpretation, we fit linear models instead of (ordered or binary) logit models for all of our dependent variables (cf. Borgatta & Bohrnstedt, 1980; Gomila, 2021). Our continuous dependent variables (trustworthiness of the Facebook post and trustworthiness of the fact check) are rescaled from -0.5 (not at all trustworthy) to $+0.5$ (extremely trustworthy) to simplify the interpretation of the model coefficients. The coding of self-reported response to the Facebook post is 0 (would not respond) and 1 (would respond). We report model results with and without covariates (frequency of Facebook use, self-reported ideology, party identification, and demographic characteristics).² Covariates are rescaled in the same manner with continuous covariates centered at their means.

Results

We find that Facebook users shown only the altered image—that is, without also getting a fact check—placed little trust in it, with a mean score of -0.17 on the -0.5 to 0.5 scale (see Figure 3; see also Table A1 in the Online Appendix contained in the replication archive). In other words, consistent with Fletcher and Park's (2017) finding of low trust in information on Facebook, we find participants had a degree of skepticism about the false post even before it was debunked.

². Given that our experiment features a negative (albeit false) report of statements by Morrison (a partisan figure), and further that we include explicitly political posttreatment variables as covariates (namely party identification and left-right ideology), we are mindful of the hazards posed by conditioning on posttreatment variables (Montgomery, Nyhan, & Torres, 2018). Though none of the treatment effects change in any meaningful way with the inclusion of covariates, we can still explicitly test whether our experimental manipulations altered respondents' self-reports of party identification and ideology by leveraging the baseline experimental condition that withheld both the Facebook post and the fact check ($n = 1,902$) excluded from the main analyses of the experiment but for which we have party identification and ideology self-reports. A linear model predicting left-right ideology as a function of the treatment group (using the baseline no Facebook post/no fact check condition as the reference group) shows no significant differences between the baseline condition and the conditions exposed to the Facebook post ($F = 1.269$, $d.f. = 5, 10131$, $p = 0.274$). We obtain the same results with binary logit models predicting Coalition party identification ($\chi^2 = 6.790$, $d.f. = 5$, $p = 0.237$) and Labor Party identification ($\chi^2 = 4.361$, $d.f. = 5$, $p = 0.499$).

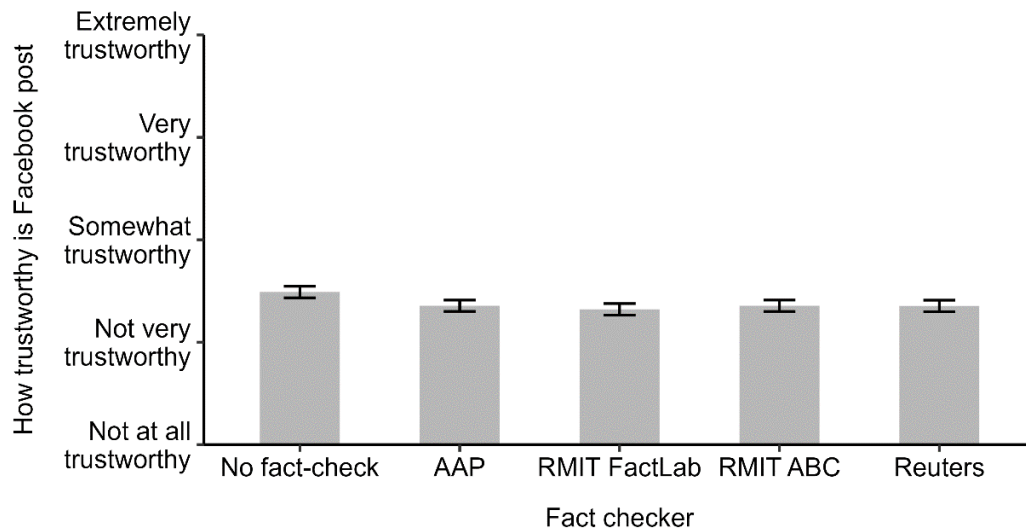


Figure 3. Perceived trustworthiness of Facebook post by experimental condition.

We further find that when the false post is paired with a fact check by any organization, Facebook users express even less trust. All of the fact-check conditions yield coefficients of approximately -0.03 to -0.04 , indicating a modest but statistically significant decrease in the trustworthiness of the altered-image Facebook post. These effects are also consistent across model specifications, with no meaningful differences whether covariates are included or excluded in the model. Thus, we find Facebook users are distrustful of the information and even more so if it is fact checked as false. In other words, fact checking “works” by disabusing Facebook users of misleading information.

Considering the political salience of the misinformation in the Facebook post—given it originally circulated during an emergency while the federal election campaign was under way—it is interesting that participants identifying with the Liberal-National Coalition express less trust in the Facebook post than those who identify with the Labor Party ($b = -0.06$, $s.e. = 0.01$, $p < 0.001$). At the same time, left-right ideology had no effect on trustworthiness. Further, we find that high frequency use of Facebook is associated with greater trust in the Facebook post, confirming Nelson and Taneja’s (2018) findings. Also, older age and higher education were correlated with reduced trust in the Facebook post (see Table A1 in the Online Appendix).

We find overall differences between the fact-checking organizations are slight, with RMIT ABC emerging as more trustworthy than the other fact-checking organizations overall (see Figure 4; see also Table A2 in the Online Appendix). This is consistent with longitudinal surveys about media trust, which regularly find ABC to be the most trustworthy news source overall (Park et al., 2022, p. 83).

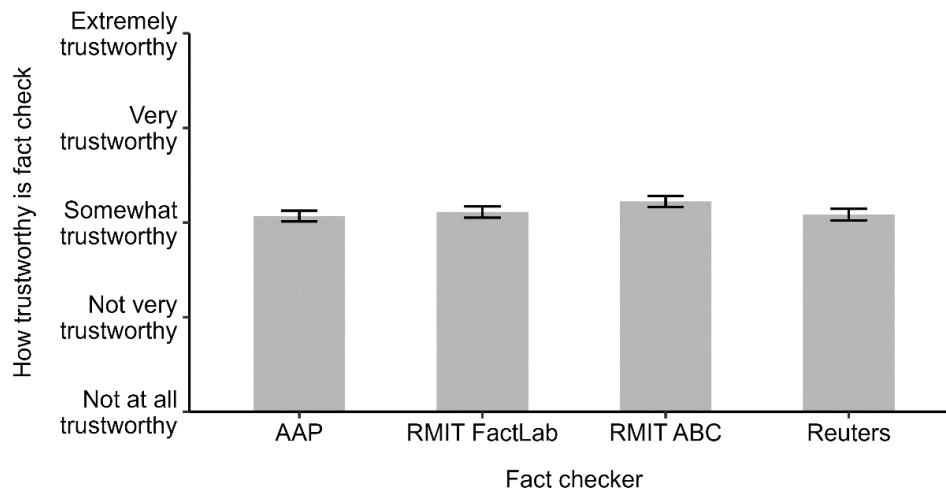


Figure 4. Perceived trustworthiness of fact-checking organizations.

Pairwise comparisons of all fact-checking organizations confirm that trust in RMIT ABC is significantly higher than AAP, RMIT FactLab, and Reuters; no other fact-checking organization is significantly different from any other. However, we find an important caveat. Those who are ideologically right-wing express less trust in fact checking ($b = -0.16$, $s.e. = 0.02$, $p < 0.001$; see Model 2.1 in Table A2 in the Online Appendix). What is more, the lower-order coefficient for ideology and the higher-order coefficient for the ideology \times RMIT ABC interaction (in Model 2.2) are negative and significant, indicating that those on the political right perceive fact checkers as less trustworthy in general, and RMIT ABC particularly so (see Figure 5). We also find a significant negative Liberal-National Coalition \times RMIT ABC interaction (in Model 2.3), indicating the Coalition supporters resemble Labor supporters in their trust in fact checkers except when it comes to the ABC (see Figure 6). These results are consistent with media trust studies cited earlier that find that individuals on the political right are the least trusting of the ABC (Jackman & Ratcliff, 2018). Just as past studies find politically right leaning voters view ABC as a more partisan news organization, the data here shows this group judge ABC-associated fact checks as less trustworthy than the other brands. The fact check perceived as most neutral by both sides of the political spectrum is the international fact checker, Reuters, which is what we would expect given it is external to Australia and its politics and has no known partisan associations.

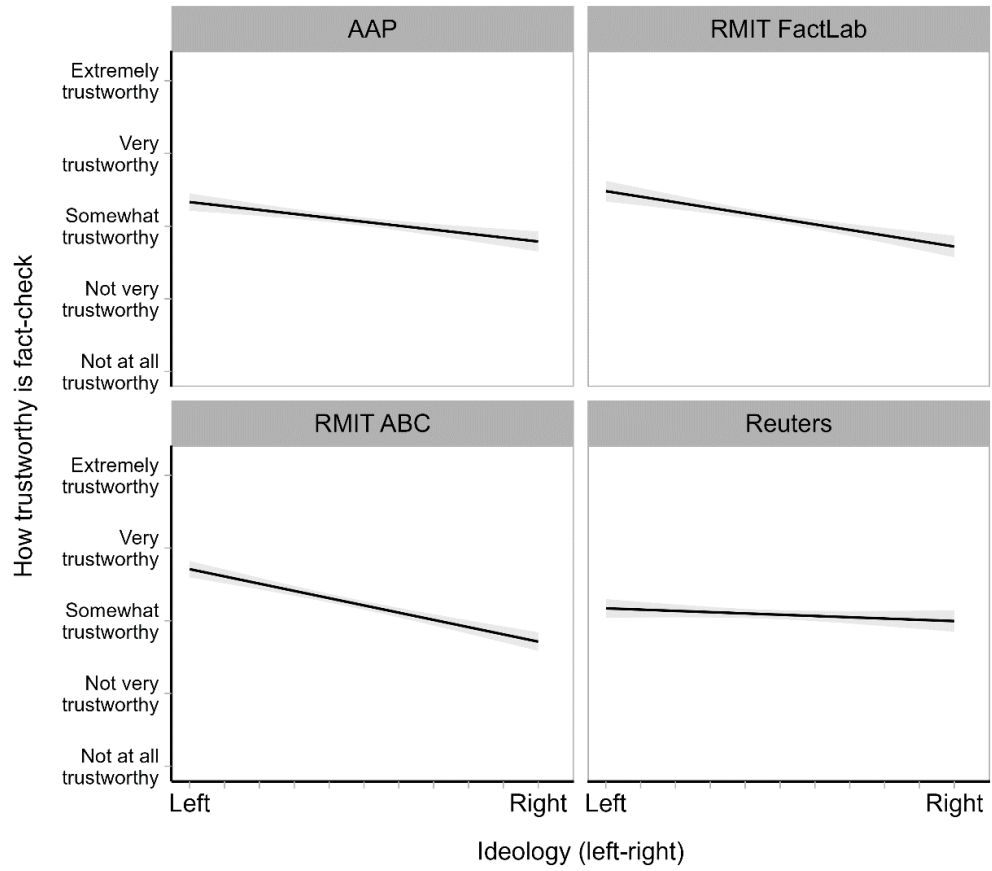


Figure 5. Left-right differences in perceived trustworthiness of fact-checking organizations.

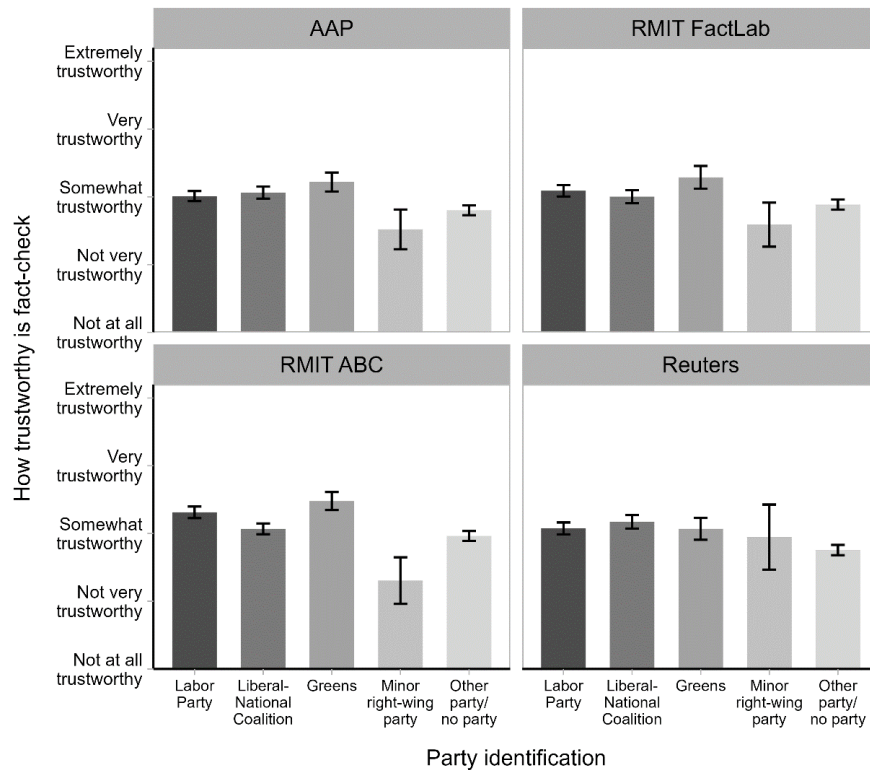


Figure 6. Partisan differences in perceived trustworthiness of fact-checking organizations.

Turning to Facebook users’ intentions to respond to the altered-image Facebook post, we find that 31.1% of Facebook users (across all experimental conditions) report an intention to react or respond to the post in some manner. The percentages for specific conditions range between 29.5% and 33.4%—differences that are not statistically significant (see Table 1).

Table 1. Intention to React to Altered-Image Facebook Post.

	No fact check (%)	AAP fact check (%)	RMIT FactLab fact check (%)	RMIT ABC Fact check (%)	Reuters fact check (%)
% who intend to react to altered Facebook post (any reaction)	33.4	30.5	30.4	29.5	31.3

Notes. χ^2 (Rao-Scott): 18.37, *d.f.* = 8, *p* = 0.368.

This nonresult is reconfirmed in a set of linear probability models (see Figure 7). While the effects of the fact-check treatments are negatively signed (and thus in the theoretically expected direction), they are not statistically significant at conventional levels (see Models 3.0 and 3.1 in Table A3 in the Online Appendix). Overall, these data offer weak support for the expectation that fact checking dampens engagement with misinformation on Facebook. By contrast, those on the political right are more likely to

react to the altered-image Facebook post in some manner, just as frequent Facebook users are (perhaps unsurprisingly) more likely to react to the post. Older Facebook users and those with higher education are less likely to react to the post.

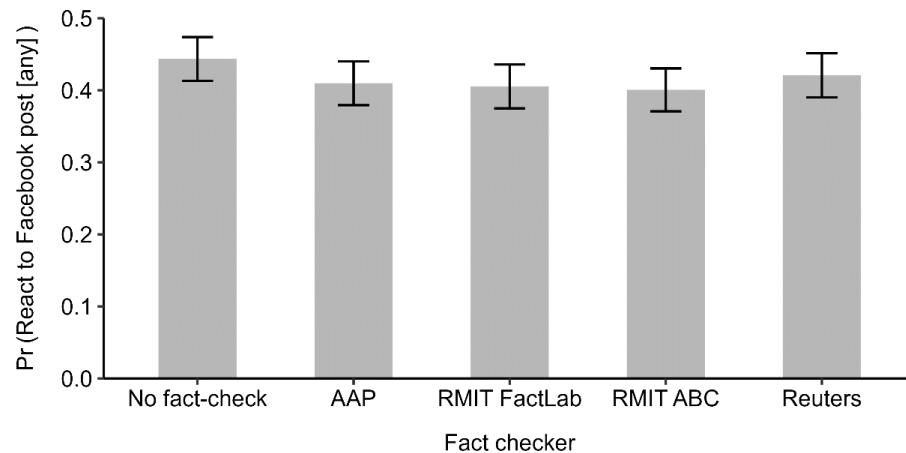


Figure 7. Intention to react to altered-image Facebook post.

Ideology features as the final predictor of engagement with the post that is worth exploration here. According to Models 3.0–3.1, those on the political right are more likely to react to the altered-image Facebook post in some manner ($b = 0.12$, $s.e. = 0.04$, $p < 0.001$). In fact, respondents generally expressed specific ideological reasons for reacting to the post in our experiment. Figure A1 (see Appendix) offers a comparison cloud of participant justifications for reacting to the post (i.e., why would you respond to the Facebook post in that way?). The figure splits those responses into three ideological categories: left, centrist, and right. The size of words in these figures represents the frequency in which they are repeated by respondents, and words closer to the center identify key language differences between the three categories. As the figure illustrates, subjects that describe themselves as ideologically left (0–4 of 10), and presumably not in support of the former conservative prime minister, largely point to Scott Morrison’s behavior as a justification for engaging with or sharing false information. For example, one ideologically left participant rationalized engaging with the post—and thereby sharing falsehoods further—stated the following:

Because although it is false it is quite likely something the pathetic prime minister of Australia would say. It fits with the narrative and anything I can do to remove this terrible man from government I shall do (within the law of course).

Conversely, those on the right claimed to be under attack from falsehoods (deliberately) spread by media—particularly Facebook—and other hostile actors, compelling some on this side of the spectrum to engage with the post to publicly confront false information. One participant stated: “I HATE the way the Left—Marxist’s lie repeatedly and the Main Stream Press and Big Tech cover for them repeatedly. Too much Left political Bias in the Press and Big Tech and ‘THEIR ABC.’”

Other respondents engaged with the false post to discredit third-party fact checking and to discredit the platform, Facebook:

I have found FaceBook [*sic*] to be the most biased organisation I've encountered. [. . .] FB tend to project the opposite of the facts in almost all instances I've witnessed. They are a profoundly dishonest and project a leftist [*sic*] agenda.

Another participant commented: "Because Facebook is run by a bunch of lunatics who use an algorithm and a woke leftist mindset to manipulate the truth. It is without doubt one of the most corrupt and despicable organisations on earth."

Centrists expressed a mix of both of these views (e.g., disliking Morrison and distrusting the platform), though one notable difference involved correcting misinformation beyond the context of a partisan battle (e.g., "Because it is rubbish, Because I don't trust it, Needs to tell the truth, Proven to be false & misleading"). Others seemed more caught in the middle: "If it was true—I would be angry. If the post was fake—I would be angry also."

Discussion and Conclusion

This study explored the effect of fact checking on trust in (false) information on Facebook. Our experiment also investigated the factors that influence trust in a fact check. Finally, we examined a fact check's potential to limit the further algorithmic spread of misleading information on Facebook via engagement with the post and use of reactions. To accomplish these tasks we designed three research questions.

In answer to *Does fact checking reduce trust in an incorrect Facebook post?* the results show there is a clear (albeit small) significant relationship between fact checking and a decrease in trust in false information on Facebook. Further, there was no meaningful difference between the fact-checker organizations in this effect. Although this overall significant reduction in trust (via fact checking) is important, we should emphasize the scope of this finding. To reiterate, this study corrected false information in a direct and concise manner, in an electoral environment where voters may already have been informed the story was false. In this context, there is very little reason for our participants to continue to trust this false information. Yet, the observation that a notable bulk of our respondents continued to express some level of trust in the falsely informed Facebook post after the treatment is concerning. Although a great deal of debate about the efficacy of fact checking exists, this particular finding points more to an effect that occurs around the edges rather than one that has the capability of correcting false information among the electorate *en masse* and is suggestive of other factors at play such as motivated reasoning.

In answer to our second research question, *What factors influence trust in a fact check?* we offer a positive overall finding about source credibility, albeit one that is moderated by political partisanship. Although it is good news for Australian fact checkers that we observe that public trust overall was high in all fact checks, it was a two-edged sword for RMIT ABC. This source significantly influenced trust in both positive and negative directions. This was not evident with other sources (e.g., Reuters, AAP) whose branding did not significantly influence levels of trust.

We find respondent ideology helps explain perceptions of trust in our model. The more politically conservative our participants were, the less likely they were to trust the fact check, and especially for RMIT ABC. This is consistent with the U.S. literature that finds conservatives are less trusting of fact checks (Robertson et al., 2020; Walter et al., 2020). It is a particularly interesting finding given that the study's fact check explicitly favored a conservative government, debunking the false claim against Scott Morrison. We would expect supporters of Morrison to trust—not distrust—a fact check that confirms the PM did not speak poorly of flood victims.

As observers of Australian politics are aware, the ABC as an institution features prominently in conservative criticisms, and subjects in this experiment expressed similar concerns (e.g., "An ABC conspiracy, [The false information] involves their ABC, I'm fed up with the character assassination of Mr. Morrison and the unrelenting personal attacks directed at him by the ABC & its adherents"). For our respondents to continue to express this skepticism of ABC despite the fact checker's favorable verdict points to an important contradictory respondent state for future researchers: That is, subjects can express distrust in false information but can also simultaneously distrust a fact check correcting that falsehood. Again, this suggests complex factors are at play in forming beliefs beyond gathering facts, such as motivated reasoning.

In relation to our final research question, *Does a fact check reduce the intention of users to react to (false) posts on Facebook?* we find no compelling evidence that fact checks dampened interactions with false information. Although we accept that many of those interactions might denounce the false claim, it is nonetheless an important finding that challenges expectations that debunking falsehoods alone will reduce their online spread. Any engagement with a Facebook post helps algorithmically further its reach, unless the platform intervenes, for example by removing the post or altering the algorithm. Without such interventions, engagement with false information on Facebook presents further opportunities to spread misinformation or disinformation online. This study finds people across the political spectrum were prepared to engage with the disinformation post, albeit for different reasons; but as the comments reveal, political partisanship and preexisting views about Facebook and the ABC were oft-cited reasons motivating their engagement. We also find high-use Facebook consumers were more willing to share content, false or otherwise, than low-use Facebook consumers. This finding about engagement strikes at the heart of fact checkers' limitations in both correcting and curbing the proliferation of online misinformation and disinformation.

To conclude, our study offers both optimism and cause for concern about fact checkers capacity to limit the spread of misinformation and disinformation online. On the positive side, we find high public trust in fact checking in Australia with little difference across brands, notwithstanding some moderating effects relating to respondents' political orientation. However, concerningly, we also find many participants are prepared to engage with disinformation despite trusting a fact check that tells them the information is false. Although we cannot be certain it is motivated reasoning alone that encourages participants to willingly share disinformation, we present qualitative and quantitative evidence that political partisanship and preexisting views about the ABC and Facebook play a role. More significantly, our study uncovers a disconnect between high trust in fact checks and their capacity to limit the spread of disinformation. This finding points to the need for platforms and stakeholders to consider more direct intervention measures beyond fact checking to limit disinformation spread.

This study is not without limitations. Of note, we apply a singular example of political disinformation and its fact check. Although we find trust in fact checking is evident across sources, we do find conservatives

are less trusting of the RMIT-ABC collaboration. Given the rise of distrust in mainstream journalism is partisan, this is noteworthy. What remains unclear, however, is whether we will find the same distrust for a nonpolitical news story containing false claims. The pandemic illuminated the danger of spreading disinformation for issues on health and vaccinations. So, there is reason to believe other news topics will incite similar distrust with serious implications for the public more broadly. We also do not test here if our findings about a falsehood and subsequent fact check on Facebook would yield the same results on other social media sites where the same Scott Morrison post was found (i.e., X—formerly Twitter, YouTube). But we can only speak for one example in one context at one time. This underscores the importance of replication across topics and platforms and, critically, across nations. Finally, heightened partisanship during an election campaign may intensify skepticism of misinformation, but also the tendency to share. This is illustrated in many of our quotes whereby election cycles create a “gloves off” moment for taking down political candidates. Regular testing of fact checking against a range of disinformation during nonpolitical times are a crucial direction for future research. Ultimately, these limitations call for replication of the experimental design across topics, platforms, and countries over time.

Our study points to a number of recommendations for digital platforms and fact checkers. Given the propensity of high-use social media consumers to engage with erroneous content, platform policies might consider targeted interventions limiting the speed through which high-use users can share false information. As emotional political content is more engaging, platforms have the opportunity to use artificial intelligence tools to target emotive keywords and track sharing during election campaigns to flag potential disinformation posts. In their selection of third-party fact checkers, platforms are understandably cautious to avoid perceptions of source bias, and, in this case, international and university fact checkers are more neutral options. However, this comes at the expense of excluding an overall well-trusted news organization such as the ABC, which is highly trusted by left-leaning voters but distrusted by conservatives. It suggests fact-checking brands might be more selectively deployed to reach different Facebook users, based on political ideology. Finally, although we find fact checking is useful to identify false claims in the first instance, and it “works” to an extent to disabuse people of disinformation, we find almost a third of users still interact with this post, which may intensify its spread without platform interventions such as removing content or slowing the algorithm. This indicates that a more expansive suite of measures in addition to fact checking is needed to further limit disinformation spread to address this pernicious global problem.

References

- Agadjanian, A., Bakhru, N., Chi, V., Greenberg, D., Hollander, B., Hurt, A., ... Woodruff, A. (2019). Counting the pinocchios: The effect of summary factchecking data on perceived accuracy and favorability of politicians. *Research & Politics*, 6(3), 1–10. doi:10.1177/2053168019870351
- Amazeen, M. A., Thorson, E., Muddiman, A., & Graves, L. (2018). Correcting political and consumer misperceptions: The effectiveness and effects of rating scale versus contextual correction formats. *Journalism & Mass Communication Quarterly*, 95(1), 28–48. doi:10.1177/1077699016678186

- Andrews, T. M., & Paquette, D. (2020, June 29). Trump retweeted a video with false covid-19 claims. One doctor in it has said demons cause illnesses. *Washington Post*. Retrieved from <https://www.washingtonpost.com/technology/2020/07/28/stella-immanuel-hydroxychloroquine-video-trump-americas-frontline-doctors/>
- Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017). Effects of editorial media bias perception and media trust on the use of traditional, citizen, and social media news. *Journalism & Mass Communication Quarterly*, 94(3), 703–724. doi:10.1177/1077699016654684
- Australian Associated Press. (2023). *AAP fact check neutrality statement and code of practice*. Australian Associated Press. Retrieved from <https://www.aap.com.au/factcheck-neutrality/>
- Australian Government. (1983). *Australian Broadcasting Corporation Act 1983*. Retrieved from <https://www.legislation.gov.au/Details/C2018C00079>
- Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., . . . Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1(2), 90–143. doi:10.1093/jssam/smt008
- Barthel, M., Mitchell, A., & Holcomb, J. (2016, December 15). *Many Americans believe fake news is sowing confusion*. Pew Research Center. Retrieved from <https://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/>
- Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of measurement: Once over again. *Sociological Methods and Research*, 9(2), 147–160. doi:10.1177/004912418000900202
- Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General*, 149(4), 746–756. doi:10.1037/xge0000673
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 114(28), 7313–7318. doi:10.1073/pnas.1618923114
- Carson, A., Gibbons, A., Martin, A., & Phillips, J. B. (2022). Does third-party fact-checking increase trust in news stories? An Australian case study using the “sports rorts” affair. *Digital Journalism*, 10(5), 801–822. doi:10.1080/21670811.2022.2031240
- Chen, J. K. T., Valliant, R. L., & Elliott, M. R. (2019). Calibrating non-probability surveys to estimated control totals using lasso, with an application to political polling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(3), 657–681. doi:10.1111/rssc.12327

- Coddington, M., Molyneux, L., & Lawrence, R. G. (2014). Fact checking the campaign: How political reporters use Twitter to set the record straight (or not). *The International Journal of Press/Politics*, 19(4), 391–409. doi:10.1177/1940161214540942
- Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences of the United States of America*, 115(49), 12441–12446. doi:10.1073/pnas.1808083115
- Dahlgren, P. (2018). Media, knowledge and trust: The deepening epistemic crisis of democracy, *Javnost—The Public*, 25(1–2), 20–27. doi:10.1080/13183222.2018.1418819
- Druckman, J. N. (2022). *Experimental thinking: A primer on social science experiments*. Cambridge, UK: Cambridge University Press.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., . . . Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. doi:10.1038/s44159-021-00006-y
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology*, 111(1), 36–54. doi:10.1111/bjop.12383
- Ellis, G. (2010). Australia and New Zealand: Two agencies on parallel tracks that sometimes diverge. In O. Boyd-Barrett (Ed.), *News agencies in the turbulent era of the Internet* (pp. 108–124). Barcelona, Spain: Generalitat de Catalunya.
- Fisher, C. (2019). Australia. In N. Newman, R. Fletcher, A. Kalogeropoulos, & R. K. Nielsen (Ed.), *Reuters institute digital news report 2019* (pp. 132–133). Oxford, UK: Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf
- Fischer, S. (2020, October 13). *Factchecking goes mainstream in Trump era*. Axios. Retrieved from <https://www.axios.com/factchecking-trump-media-baad50cc-a13f-4b73-a52a-4cd9e63bd2fc.html>
- Fletcher, R., & Nielsen, R. (2019). Generalised scepticism: How people navigate news on social media. *Information, Communication & Society*, 22(12), 1751–1769.
- Fletcher, R., & Park, S. (2017). The impact of trust in the news media on online news consumption and participation. *Digital Journalism*, 5(10), 1281–1299. doi:10.1080/21670811.2017.1279979
- Garrett, R. K., Nisbet, E. C., & Lynch, E. K. (2013). Undermining the corrective effects of media-based political factchecking? The role of contextual cues and naïve theory. *Journal of Communication*, 63(4), 617–637. doi:10.1111/jcom.12038

- Gibbons, A., & Carson, A. (2022). What is misinformation and disinformation? Understanding multi-stakeholders' perspectives in the Asia Pacific. *Australian Journal of Political Science*, 57(3), 231–247. doi:10.1080/10361146.2022.2122776
- Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, 150(4), 700–709. doi:10.1037/xge0000920
- Gravelle, T. B. (2021). The measurement invariance of customer loyalty and customer experience across firms, industries, and countries. *Methods, Data, Analyses*, 15(2), 191–214.
- Gravelle, T. B., Phillips, J. B., Reifler, J., & Scotto, T. J. (2022). Estimating the size of “anti-vax” and vaccine hesitant populations in the U.S., U.K., and Canada: Comparative latent class modelling of vaccine attitudes. *Human Vaccines & Immunotherapeutics*, 18(1), 1–10. doi:10.1080/21645515.2021.2008214
- Graves, L. (2016). *Deciding what's true: The rise of political factchecking in American journalism*. New York, NY: Columbia University Press.
- Graves, L. (2017). Anatomy of a factcheck: Objective practice and the contested epistemology of factchecking. *Communication, Culture & Critique*, 10(3), 518–537. doi:10.1111/cccr.12163
- Jackman, S., & Ratcliff, S. (2018, February 18). *America's trust deficit*. United States Studies Centre, University of Sydney. Retrieved from <https://www.usssc.edu.au/analysis/americas-trust-deficit>
- Jarman, J. W. (2016). Influence of political affiliation and criticism on the effectiveness of political factchecking. *Communication Research Reports*, 33(1), 9–15. doi:10.1080/08824096.2015.1117436
- Krupnikov, Y., & Levine, A. S. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, 1(1), 59–80. doi:10.1017/xps.2014.7
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. doi:10.1126/science.aao2998
- Lenz, G. S. (2012). *Follow the leader? How voters respond to politicians' policies and performance*. Chicago, IL: The University of Chicago Press.
- Meese, J., & Hurcombe, E. (2020, November 12). *Regulating misinformation: Policy brief*. RMIT University. Retrieved from <https://apo.org.au/node/309357>

- Miragliotta, N., & Errington, W. (2012). The rise and fall and rise again of public broadcasting? The case of the Australian Broadcasting Corporation. *Australian Journal of Public Administration*, 71(1), 55–64. doi:10.1111/j.1467-8500.2012.00755.x
- Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760–775. doi:10.1111/ajps.12357
- Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2), 109–138. doi:10.1017/XPS.2015.19
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton, NJ: Princeton University Press.
- Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. *New Media & Society*, 20(10), 3720–3737. doi:10.1177/1461444818758715
- Nieminen, S., & Rapeli, L. (2019). Fighting misperceptions and doubting journalists' objectivity: A review of factchecking literature. *Political Studies Review*, 17(3), 296–309. doi:10.1177/1478929918786852
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. doi:10.1007/s11109-010-9112-2
- Nyhan, B., Reifler, J., & Ubel, P. A. (2013). The hazards of correcting myths about health care reform. *Medical Care*, 51(2), 127–132. doi:10.1097/MLR.0b013e318279486b
- Owen, T., Loewen, P., Ruths, D., Bridgman, A., Gorwa, R., MacLellan, S., . . . Zhilin, O. (2019, October). *Digital democracy project: Research memo# 5-fact-checking, blackface and the media*. Public Policy Forum. Retrieved from <https://apo.org.au/node/263381>
- Park, S., Fisher, C., Flew, T., & Dulleck, U. (2020). Global mistrust in news: The impact of social media on trust. *International Journal on Media Management*, 22(2), 83–96. doi:10.1080/14241277.2020.1799794
- Park, S., McCallum, K., Lee, J., Holland, K., McGuinness, K., Fisher, C., & John, E. (2022). *Covid-19: Australian news & misinformation longitudinal study*. Canberra: News & Media Research Centre. Retrieved from <https://www.canberra.edu.au/research/faculty-research-centres/nmrc/research/covid-19-australian-news-and-misinformation>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. doi:10.1037/xge0000465
- Pingree, R. J., Hill, M., & McLeod, D. M. (2013). Distinguishing effects of game framing and journalistic adjudication on cynicism and epistemic political efficacy. *Communication Research*, 40(2), 193–214. doi:10.1177/0093650212439205

- Ren, Z., Dimant, E., & Schweitzer, M. (2023). Beyond belief: How social engagement motives influence the spread of conspiracy theories. *Journal of Experimental Social Psychology, 104*(3), 1–12. doi:10.1016/j.jesp.2022.104421
- RMIT. (2023). *RMIT ABC Fact Check*. Retrieved from <https://www.rmit.edu.au/about/schools-colleges/media-and-communication/research/public-communication-research-and-advisory-services/projects/fact-check>
- RMIT FactLab. (2022, March 29). *Prime Minister did not say Queensland flood victims should be "grateful for government assistance."* RMIT FactLab. Retrieved from <https://www.rmit.edu.au/news/factlab-meta/queensland-flood>
- Robertson, C. T., Mourão, R. R., & Thorson, E. (2020). Who uses factchecking sites? The impact of demographics, political antecedents, and media use on factchecking site awareness, attitudes, and behavior. *The International Journal of Press/Politics, 25*(2), 217–237. doi:10.1177/1940161219898055
- Robinson, S. (2018). Crisis of shared public discourses: Journalism and how it all begins and ends with trust. *Journalism, 20*(1), 56–59. doi:10.1177/1464884918808958
- Shepherd, K. (2020, March 24). A man thought aquarium cleaner with the same name as the anti-viral drug Chloroquine would prevent coronavirus. It killed him. *Washington Post*. Retrieved from <https://www.washingtonpost.com/nation/2020/03/24/coronavirus-chloroquine-poisoning-death/>
- Shin, J., & Thorson, K. (2017). Partisan selective sharing: The biased diffusion of factchecking messages on social media. *Journal of Communication, 67*(2), 233–255. doi:10.1111/jcom.12284
- Simis, M. J., Madden, H., Cacciatore, M. A., & Yeo, S. K. (2016). The lure of rationality: why does the deficit model persist in science communication? *Public Understanding of Science, 25*(4), 400–414. doi:10.1177/0963662516629749
- Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition, 9*(3), 286–299. doi:10.1016/j.jarmac.2020.06.006
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755–769. doi:10.1111/j.1540-5907.2006.00214.x
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication, 37*(3), 350–375. doi:10.1080/10584609.2019.1668894
- Williams, I., Gravelle, T. B., & Klar, S. (2022). The competing influence of policy content and political cues: Cross-border evidence from the United States and Canada. *American Political Science Review, 116*(4), 1375–1388. doi:10.1017/S0003055422000351

Appendix



Figure A1. Subject justifications for reacting to the post.