The Corrections Dilemma: Media Retractions Increase Belief Accuracy But Decrease Trust

Joshua Freitag¹, Madeline Gochee¹, Mitchell Ransden¹, Brendan Nyhan¹*, Kristy Roschke² and Dan Gillmor²

¹Dartmouth College, Hanover, NH 03755, USA and ²Arizona State University, Tempe, AZ 85281, USA

*Corresponding author. Email: nyhan@dartmouth.edu

Abstract

Why are prominent news media retractions so rare? Using data from a survey experiment in which respondents view simulated Twitter newsfeeds, we demonstrate the dilemma facing news organizations that have published false information. Encouragingly, media retractions are effective at informing the public – they increase the accuracy of news consumers’ beliefs about the retracted reporting more than information from third parties questioning the original reporting or even the combination of the two. However, trust in the news outlet declines after a retraction, though this effect is small both substantively and in standardized terms relative to the increase in belief accuracy. This reputational damage persists even if the outlet issues a retraction before a third party questions the story. In a social media environment that frequently subjects reporting to intense scrutiny, the journalistic mission of news organizations to inform the public will increasingly conflict with organizational incentives to avoid admitting error.

Keywords: Correction; media; news; retraction; trust

Introduction

News reporting inevitably contains errors (e.g., Maier, 2005), creating a dilemma for journalists who seek to uphold their professional responsibilities (Bugeja, 2007; Kampf and Daskal, 2014). Issuing a retraction or correction may increase the accuracy of the public’s beliefs about the story in question, but it can also harm trust in the organization, making the public less likely to believe or consume their reporting.
in the future – an especially damaging prospect at a time when trust in the media is low and industry business models are under threat (Abernathy, 2018; Brenan, 2021).

The challenge that media organizations face is particularly acute on social media platforms, an important news source for about half of Americans (Walker and Matsa, 2021). Social media can quickly spread false news reports, increasing the need for rapid corrections from the organizations that published them. However, such corrections threaten to further harm trust in the media outlet in question if widely publicized. It is therefore essential to determine the effect of retractions and corrections on both public beliefs and trust in the news outlets that issue them.

Importantly, corrections of specific errors or retractions of incorrect stories are likely to be successful. Recent studies indicate backfire effects are extremely rare (e.g., Porter and Wood, 2019; Swire-Thompson et al., 2020). In general, experimental studies find that corrective information is typically effective at reducing belief in false claims even about controversial topics (Nyhan et al., 2019; Walter and Murphy, 2018). As a result, journalistic corrections should increase the accuracy of beliefs held by people who see them, at least partially offsetting the effect of misinformation exposure.

However, learning about errors made by news organizations can potentially undermine trust. Observational survey data indicate that perceptions of news inaccuracy are correlated with reduced trust in the media (Brosius et al., 2022; Wilner et al., 2022). Similarly, people say they trust the media less when it makes errors (Gronke and Cook, 2007). Though the public claims to value media organizations that admit mistakes and says that it views errors that are corrected somewhat more favorably (Van der Wurff and Schönbach, 2014; Karlsson et al., 2017), they report no greater trust in news that includes a correction in the one experiment that has been conducted to date (Karlsson et al., 2014). However, that study was conducted in Sweden, it only tested the presence of a minor correction (an error in the cost of a water park), and it did not consider the effect of learning from a third party about an error that a media organization does not itself disclose.

We examine these questions in the context of a simulated social media environment, allowing us to examine how contemporary news consumers react to admissions of error by media organizations. This emerging methodology attempts to better capture the real-time social media experience by using experimental stimuli that are embedded in mock newsfeeds (e.g., Jahng and Littau, 2016; Vraga and Bode, 2017; Kaiser, Keller, and Kleinen-von Königslöw, 2021). Our study applies this method to the study of media retractions for the first time.

Specifically, we conduct an experiment in which participants twice view a mock Twitter feed. In the first set of tweets, respondents viewed a tweet from a news organization (the Canadian Broadcasting Corporation [CBC]) about an ISIS story they published. In the second set of tweets, respondents were randomized to see a tweet by the news organization retracting their ISIS story, a tweet by a third party questioning the story, both (in random order), or neither. The treatments seek to compare the effect of exposure to a correction from a news organization with learning about the errors from a third party.

We predicted that a correction by a news organization or information questioning the report from another source would each increase belief accuracy but decrease trust in the organization. In real life, however, both may occur. We therefore also measured the effects of exposure to both messages on belief accuracy and news outlet trust.
Our results confirm the existence of what we call the corrections dilemma – news outlets that disclose their mistakes improve the accuracy of readers’ beliefs, but reduce trust in their organization. While the negative effects of a retraction on news outlet trust are substantially smaller in standardized terms, these competing effects on accuracy and trust may pose challenges for journalists and editorial teams when deciding how to respond to a story that contains errors. We further find that third party messages questioning news reporting lead to similar, but smaller, effects. When participants encountered both types of corrections, the effects on belief accuracy were actually smaller compared to a correction from only the news outlet, but the decrease in trust was similar. These effects were moderated by partisan affiliation (positive effects on belief accuracy were smaller among Republicans) and overall trust in the mass media (negative effects on trust were greater among those with higher trust) but not Twitter use.

**Theoretical expectations**

Based on the research described above, we offer the following preregistered hypotheses and research questions.\(^1\) The first set estimates the effects of exposure to various messages about the initial news report in isolation. First, we hypothesized that exposure to a media retraction (H1a) or information questioning the report from a third party (H1b) would increase the accuracy of respondent’s beliefs by increasing doubt in the initial report. We also hypothesized that exposure to a third party message questioning the initial report would reduce the perceived trustworthiness of the media outlet that published it (H2). By contrast, we have less prior evidence about the effects of corrections and retractions on media outlet trust. We therefore preregistered a research question asking if retracting the story would affect trust in the retracting outlet (RQ1).

We also consider the joint effects of exposure to a news outlet retraction and a third party message questioning the initial report. We expected that effects on belief accuracy would be greatest when both the news outlet and third party question the initial report (H1c) due to the repetition of the message from multiple sources. We also sought to investigate whether the effects of exposure to a third party message questioning the original reporting vary by whether the news outlet is willing to admit error. We therefore asked if the effects of a third party questioning a false report on news outlet trust would vary by whether the news organization corrects the report itself (RQ2a) and if those effects depend on whether people learn of the error from the outlet first (RQ2b).

**Methods**

**Sample characteristics**

We recruited a survey sample using Amazon’s Mechanical Turk, which has been shown to generate valid data (Berinsky et al., 2012) and to mirror estimates of causal effects from representative samples (Coppock, 2019). To further improve the quality

---

\(^1\) The preregistration is available at [https://osf.io/yxhme/](https://osf.io/yxhme/).
of our sample, we restrict our sample to CloudResearch-approved participants. This service screens participants from the platform for attention and engagement, overcoming limitations of basic Turk sampling (Litman et al., 2017). We also apply CloudResearch tools to block submissions with duplicate or suspicious IP addresses and workers whose IP addresses place them outside the United States.

Following our preregistration, we collected 2620 responses in our first wave and then recruited an additional 500 respondents who self-identify as Republican to achieve approximate partisan balance. We excluded data from 258 participants who failed a pre-treatment attention check (see Online Appendix A for survey items). Of the remaining 2,862 respondents, 45% were male, 56% were college graduates, 74% were white, and 66% were 25–44 years old. 38% identified as Democrat and 9% leaned Democrat; 36% identified as Republican and 8% leaned Republican.

**Experimental design**

We conducted a between-subjects experiment following the design outlined in Figure 1. Respondents answer a set of baseline measures, view a simulated Twitter feed including a false news report, answer questions related to its content including belief in the claim in question and trust in the outlet that reported it, and then are randomized to see a retraction from the media organization, a message from a third party questioning the initial report, both (in random order), or neither. We then again measure belief in the claim in question and trust in the outlet that reported it post-treatment. We summarize the full procedure below.

After answering a standard battery of demographic and attitudinal questions and passing two attention checks (see Online Appendix A), all respondents viewed a simulated Twitter feed that included a modified version of a real tweet from the Canadian Broadcast Corporation (CBC) stating that “In 2014, a jihadi called Abu Huzaifa al-Kanadi left Toronto to become an ISIS executioner. Five months later, he decided to escape.” This stimuli was inspired by The New York Times podcast “Caliphate,” which initially publicized such a claim but later had to retract key
elements of the show after the credibility of their source, a man named Shehroze Chaudry, was called into question (Folkenflik, 2020). Chaudry had also spoken to the CBC and gave it a story that differed in part from what he said on “Caliphate” (Swain, 2018). Both the CBC and the Times retracted their stories on Chaudry’s claims, allowing us to measure the effects of the real-world process of journalistic correction. We chose to attribute the story in the experiment to the CBC, a less well-known and less polarizing news outlet among Americans, to avoid source effects related to pre-existing attitudes toward the Times.

After answering questions measuring their factual beliefs (including in the claim that a Canadian had become an ISIS executioner) and trust in news outlets (including the CBC), participants were then randomly assigned with equal probability using simple randomization in Qualtrics to one of five conditions in which they viewed a second set of tweets in random order:

- No correction: Respondents do not encounter any further information about the story
- News outlet correction: Respondents receive a tweet from the news outlet retracting the story ("reporting... fell short of our standards")
- Third party correction: Respondents receive a tweet from a third party identifying the story as false ("Who did he lie to, and why did both organizations publish such an unreliable source?")
- Both corrections (news outlet first): Respondents receive a tweet from the news outlet retracting the story followed by a tweet from a third party questioning the story
- Both corrections (third party first): Respondents receive a tweet from a third party questioning the story followed by a tweet from the news outlet retracting the story

The stimuli used in the study were based on actual messages about the “Caliphate” podcast but adapted for clarity. Engagement with the initial and corrective tweets (i.e., likes and retweets) is slightly varied but is kept similar to minimize confounding. Notably, the modified CBC retraction tweet directly admits that the initial reporting is unreliable, while the third party message only questions its dependence on a questionable source. The latter type of criticism is typical of the media criticism that news consumers often encounter on social media platforms, which often takes the form of “just asking questions” since third parties can rarely prove that a story is false. (We discuss differences between the retraction and the third party message questioning the initial report further in our conclusion.)

The other tweets in the study were created based on neutral tweets identified by Vraga et al. (2016). All filler tweets were created with false names and stock images and were attributed to a mix of verified individuals, unverified individuals, and verified news organizations to resemble a real Twitter feed. While the tweets mostly mirror the content of the original neutral tweets from Vraga and Bode (2017), modifications were made to coincide with real news stories around the time of the tweets (2019 and 2020). All dates are standardized within the mock feeds and tweet reactions are randomized. All stimuli and question wording are provided in Online Appendix A.
Outcome measures

We consider two primary outcome variables in our analyses. “Belief accuracy” refers to the accuracy of respondents’ belief in the claim that “A Canadian became an ISIS executioner.” We measure this outcome on a four-point scale from “Very accurate” (1) to “Not at all accurate” (4) with higher values indicating greater accuracy. “News outlet trust” measures the amount of trust and confidence that respondents have that the CBC will report the news fully, accurately, and fairly on a four-point scale from “None at all” (1) to “Great deal” (4). All outcome variables are measured before and after the experimental treatment.

Analytical strategy

We estimate treatment effects using OLS regressions with robust standard errors using listwise deletion in cases of item nonresponse (2.8% or 79 responses for belief accuracy; 2.7% or 76 responses for news outlet trust). Control variables were selected for each outcome variable using Least Absolute Shrinkage and Selection Operator to maximize precision (Bloniarz et al., 2016) from the following set of variables:

- Pre-treatment measures of outcome variables
- Education (college graduate indicator) (factor variable)
- Age group (18–24, 25–34, 35–44, 45–54, 55–64, 65+) (factor variable)
- Male (1/0) (factor variable)
- Political interest (five-point)
- Non-Hispanic white (1/0) (factor variable)

All analyses follow our preregistration unless otherwise specified (https://osf.io/yxhme/).

Results

To verify that our study would not be contaminated by prior knowledge about the “Caliphate” podcast (Druckman and Leeper, 2012), we fielded a smaller, pre-experiment survey measuring baseline knowledge about the controversy among the same population of CloudResearch-approved participants on Amazon Mechanical Turk (question wording is provided in Online Appendix A). In total, only six of the 253 respondents who participated and passed our attention checks (2.4%) could identify the New York Times as a major outlet that had issued a high-profile retraction in the last year and then identify the topic as “ISIS whistleblower.” Only one could name the “Caliphate” podcast as the source of the retraction in an open-ended follow-up. We therefore infer that our experimental sample (which excluded respondents to this survey) had very low levels of prior knowledge about the retraction in question.

We therefore turn to data from the main study and consider the effects of our experimental treatments on the accuracy of people’s beliefs about the ISIS whistleblower story and their trust in the CBC. The results, which are summarized in Table 1, provide support for our hypotheses that a media outlet retraction or a
message from a third party questioning an initial report would decrease belief in its central claim. Specifically, we find that a retraction from the outlet that published the false story increased the accuracy of respondents’ beliefs by 0.86 points on a four-point scale (95% CI: 0.77–0.94, p < 0.005). Corrections from a third party source were much less effective, increasing belief accuracy by 0.16 points (95% CI: 0.07–0.24, p < 0.005). These effects were substantively significant. Belief that the claim that “A Canadian was an ISIS executioner” was either “somewhat” or

---

2We provide details on manipulation checks testing respondent recognition of the tweets they have seen in Online Appendix B.

---

Table 1.
Main Effects on Belief Accuracy and News Outlet Trust

<table>
<thead>
<tr>
<th></th>
<th>Belief accuracy</th>
<th>News outlet trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>News outlet retraction</td>
<td>0.86***</td>
<td>−0.11***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Third party questioning</td>
<td>0.16***</td>
<td>−0.06*</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Both messages (news outlet first)</td>
<td>0.68***</td>
<td>−0.17***</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Both messages (third party first)</td>
<td>0.57***</td>
<td>−0.14***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

Control variables

✓✓

Differences in treatment effects

<table>
<thead>
<tr>
<th></th>
<th>Belief accuracy</th>
<th>News outlet trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both messages (news outlet first) – news outlet correction</td>
<td>0.179***</td>
<td>−0.057*</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Both messages (news outlet first) – third party correction</td>
<td>−0.520***</td>
<td>−0.105***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Both messages (third party first) – news outlet correction</td>
<td>0.287***</td>
<td>−0.030*</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Both messages (third party first) – third party correction</td>
<td>−0.412***</td>
<td>−0.078**</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Both messages (news outlet first) – both messages (third party first)</td>
<td>−0.108*</td>
<td>−0.028*</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.029)</td>
</tr>
</tbody>
</table>

N 2745 2747

OLS regression with robust standard errors; * p<0.05, ** p<0.01, *** p<0.005 (two-sided). Belief accuracy and news outlet trust were measured on four-point scales (see Online Appendix A): Belief accuracy controls: pre-treatment levels of belief accuracy, age 65 or older; News outlet trust controls: political interest, race, age 18–24 or 55–64, baseline news outlet trust, pre-treatment news outlet trust.

https://doi.org/10.1017/XPS.2023.4 Published online by Cambridge University Press
“very” accurate decreased from 63.3% in the control condition to 53.2% in the third party condition and 19.0% in the news outlet retraction condition.

As expected, viewing both messages increased participant belief accuracy more than the third party message questioning the initial report alone. Contrary to our expectations, however, the combined effect of exposure to both the news outlet and third party messages on belief accuracy was less than the news outlet correction alone regardless of order (news outlet first: 0.68, 95% CI: 0.59–0.76, p < .005; third party first: 0.57, 95% CI: 0.48–0.66, p < .005). Overall, these results are consistent with past findings demonstrating the effectiveness of corrective information at increasing belief accuracy but suggest that the presence of the third party message diluted the effect of the media retraction or confused readers.

Despite increasing the accuracy of respondents’ beliefs, the treatment messages casting doubt on the reliability of the initial story also reduced trust in the news outlet that published it. The news outlet retraction decreased trust in the source of the story by 0.11 points on a four-point scale (95% CI: 0.06–0.11, p < 0.005). Third party corrections caused a similar decrease in trust of 0.06 points (95% CI: 0.02–0.11, p < 0.05). Exposure to both messages had slightly larger negative effects, decreasing trust by 0.17 points when the retraction from the original source was viewed first (95% CI: 0.12–0.22, p < 0.005) and by 0.14 points when the message from a third party was viewed first (95% CI: 0.09–0.19, p < 0.005). The order in which these messages were viewed had no measurable effect on trust. However, the joint effects of seeing two messages questioning the initial report were significantly more negative than seeing either message alone in three of four cases (see Table 1).

Figure 2 visualizes the treatment effect estimates from Table 1 to highlight the trade-off between belief accuracy and trust in the news outlet in question.

Specifically, all four treatments significantly increase belief accuracy but decrease trust in the news outlet that published the initial reporting. However, we note that the effects of the retraction on belief accuracy are much larger in standardized terms than the negative effect on trust (Cohen’s d of 0.91 vs. 0.12). The standardized effect on accuracy is also somewhat larger for the message from a third party questioning the initial report, though the magnitudes of each are smaller (d = 0.17 for accuracy, d = 0.07 for trust).

We explore three potential moderators of these effects as part of a preregistered research question. First, we find that party affiliation often moderated treatment effects on belief accuracy, which were generally smaller in magnitude (i.e., less positive) for Republicans than for Democrats and independents. However, evidence of partisan differences in effects on news outlet trust is less consistent. Second, we find no evidence that our results are moderated by Twitter use. Finally, we find little evidence of differences in belief accuracy depending on levels of media trust, but find that corrections reduce news outlet trust most among respondents with the highest levels of media trust in general. Details of these analyses are reported in Online Appendix B.

Finally, we also examine the preregistered research question of whether the treatments affected people’s epistemic political efficacy – that is, their confidence in their ability to perceive reality surrounding political issues (Pingree, et al., 2014). We found no significant treatment effects on epistemic political efficacy. These results are reported in greater detail in Online Appendix B.
Conclusion

Our results indicate that when news organizations realize that they have published a factually inaccurate story, they face a potential dilemma when choosing how to respond. Retracting their story will better inform their audience, but reduce trust in their news organization. However, these concerns may be mitigated to some extent by the relative size of these effects. Though observers and journalists may weigh these effects differently, we find that the negative effects on trust are small both substantively and in standardized terms relative to the informational benefit provided to their audience. Retracting a story increases respondent belief accuracy dramatically but only modestly decreases trust in the news outlet. If a third party has also questioned the story, the marginal effects of a retraction are similar (a substantial additional increase in belief accuracy and a modest additional decrease in trust).

Of course, our study has important limitations. Most notably, we considered a single, low-salience issue on Twitter with an unfamiliar news organization and third party critic. Varying one or more of these elements may change the observed effects on belief accuracy and news outlet trust or their relationship with potential moderators. For instance, studies that consider partisan issues or more prominent news organizations might find wider partisan differences. Additional research should similarly test the effects of retractions on trust in more prominent news outlets,
which might suffer less reputational damage in response to a correction. Future studies might also investigate the effect of varying the social media platform tested or altering how the newsfeed environment is represented. In addition, further research should seek to validate our findings using a nationally representative sample and test whether our results apply outside the USA. Fourth, future studies could vary the messages presented by both the media outlet (e.g., a correction rather than a retraction) and the third party (e.g., flatly stating the story is false rather than raising questions). Finally, it would be valuable to measure the long-term effects of the retraction on trust in the news outlet and consumption of its content.

Ultimately, this study helps us understand the incentives news organizations face to avoid admitting error. However, our findings also demonstrate that efforts to set the record straight can have substantial effects on the accuracy of people’s beliefs. Future research should thus seek to determine how to minimize the costs of publicly admitting error, which could increase the incentives for media outlets to publish more frequent and prominent corrections and in turn help contribute to a better-informed public.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/XPS.2023.4.

Data availability. The data and code required to replicate all analyses in this article are available in the Journal of Experimental Political Science Dataverse at https://doi.org/10.7910/DVN/42UB2N (Freitag et al., 2023).

Acknowledgments. We thank the Arizona State University News Co/Lab, Craig Newmark Philanthropies, Dartmouth College Undergraduate Advising & Research, and the Paul K. Richter and Evalyn E. Cook Richter Memorial Fund for generous funding support.

Conflicts of interest. The authors declare no conflicts of interest.

Ethics statement. This research was approved by the Dartmouth Committee for the Protection of Human Subjects (STUDY00032361) and adheres to APSA’s Principles and Guidelines for Human Subjects Research. Further details are provided in Online Appendix C.

References


