Cognition 258 (2025) 106090

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Belief updating in the face of misinformation: The role of source reliability

Greta Arancia Sanna *, David Lagnado

Department of Experimental Psychology, University College London, UK

ARTICLE INFO

Keywords:
Belief updating
Source reliability
Misinformation
Reasoning
Cognitive processes
Trustworthiness
Expertise

ABSTRACT

This paper investigates the process of belief updating in the presence of contradictory and potentially misleading information, focusing on the impact of source reliability. Across four experiments, we examined how individuals revise their beliefs when confronted with retracted information and varying source credibility. Experiment 1 revealed that participants discounted retracted information and reverted to their prior beliefs, in contrast to the Continued Influence Effect commonly reported in the literature. Experiment 2 demonstrated that source reliability significantly influences belief updating: reliable sources led participants to discount initial allegations more effectively than unreliable sources. Experiments 3 and 4 examined how people update their beliefs given opposing sources of differing reliability; we found that participants appropriately incorporated source reliability and penalised sources that were corrected, regardless of the corrector's reliability. Additionally, in contrast to previous research, both trustworthiness and expertise contributed to judgments of source reliability. Our results resolve some of the mixed findings in previous research, and highlight that individuals' belief updating are rationally sensitive to differences in source reliability. Our findings have broad implications for correcting misinformation in political, medical, and other applied contexts, and further underscore the need to ground misinformation correction strategies in robust psychological research.

1. Introduction

The spread of misinformation has been identified by the World Economic Forum as one of the top ten perils to society (World Economic Forum, 2024). A major consequence of misinformation is its potential to influence voter behaviour and decision-making, a concern highlighted during pivotal events like the UK Brexit vote and the 2016 US elections, where false and misleading information was widely disseminated, potentially shaping public opinion (Bastos & Mercea, 2019; Oyserman & Dawson, 2020; Ross & Rivers, 2018). In the 2024 US elections Vice President Kamala Harris was targeted by fake news including one claim that she was involved in a hit-and-run (Guardian, 2024). These types of misinformation not only risk eroding public trust but also intensify political polarisation, reinforcing ideological divides and undermining democratic processes (e.g., Ribeiro et al., 2017; Waldman, 2017).

The consequences of such claims are exacerbated by the fact that even clear and credible corrections can fail to eliminate the effect of misinformation (Anderson et al., 1980; Guillory & Geraci, 2010; Johnson & Seifert, 1994; Ross et al., 1975; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999). In an early set of studies Anderson et al. (1980) found that individuals persevered in their belief about a claim - e.g. risk-taking is (not) conducive to success in firefighting - even after the only evidence provided in its support was discredited. Similarly, Ross et al. (1975) showed that participants given feedback on their performance in a task persevered in maintaining a self-perception consistent with that feedback even after it was identified as randomly assigned. Note that these findings could be explained by either normative or non-normative reasons for why people fail to update their beliefs in face of new evidence. For example, although Anderson et al. claim that individuals "cling to their beliefs to a considerably greater extent than is logically or normatively warranted" (Anderson et al., 1980, p. 1045) they also concede the possibility that individuals might not have believed in the discredited information. Similarly, Ross et al. (1975) suggest that feedback may prompt individuals to search for additional evidence supporting the initial claim. Once the initial evidence for the claim is discredited, the additional evidence still reinforces the belief, making it reasonable for individuals not to revise their stance.

This phenomenon – now termed the *Continued Influence Effect* (CIE) - has been tested in numerous studies. In the classic version, for example, participants were told that a fire had developed at a commercial warehouse (Guillory & Geraci, 2010; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999). An initial message

* Corresponding author at: University College London, 26 Bedford Way, London WC1H 0AP, UK.
E-mail address: greta.sanna.23@ucl.ac.uk (G.A. Sanna).

https://doi.org/10.1016/j.cognition.2025.106090
Received 23 July 2024; Received in revised form 14 February 2025; Accepted 16 February 2025
Available online 21 February 2025

stated that the storage room contained carelessly stored paint cans and gas containers. Later, participants read a correction stating that this information was inaccurate and that the storage room was empty. Despite this correction, participants often disregarded the updated information when making inferences about the story. For instance, when asked about the possible cause of the fire, many participants still suggested that it could have been started by the flammable materials left in the storage room (Wilkes & Leatherbarrow, 1988). Thus, once people encountered critical information they kept relying on it, even after it had been retracted. Importantly, participants were able to recall the correction when specifically prompted and they accurately answered factual questions. However, when responding to inference-based questions, they reverted to using the original, retracted information. While memory may still contribute to this process, it seems likely that cognitive processes beyond memory—potentially related to inference and belief updating—also play a significant role. These cognitive factors, in conjunction with memory processes, may interact to influence belief persistence when confronted with contradictory information. In sum, misinformation that is initially presented as true but is later identified as false is shown to have an ongoing influence on inferential reasoning.

The Continued Influence Effect has been extensively studied (Ecker & Antonio, 2021; Chan et al., 2017; Johnson & Seifert, 1994; Lewandowsky et al., 2012; Paynter et al., 2019; Walter & Murphy, 2018; Walter & Tukachinsky, 2020; Wilkes & Leatherbarrow, 1988). Most studies have found that retractions are partially successful at reducing reliance on misinformation but do not completely eliminate its influence back to baseline (e.g., Ecker et al., 2011; Ecker et al., 2017; Ecker & Antonio, 2021).

One line of research has looked at the CIE in impression formation, that is the processes by which different pieces of knowledge about another person are combined into a global or summary impression (Uleman & Kressel, 2013). Here the findings are mixed. Some studies suggest that misinformation continues to influence a person's impression: for example, when inadmissible evidence still affects jury decisions (Steblay et al., 2006) or when rumours affect voter preferences (Jardina & Traugott, 2019; Nyhan & Reifler, 2010; Weeks & Garrett, 2014). One relevant study by Thorson (2016), found that misinformation about a fictional politician accepting donations from a convicted felon negatively affected the candidate's evaluation even after the information was retracted. The significant findings, however, were limited to the conditions where the misinformation in the study was worldview congruent, namely where the political candidate was identified as belonging to the political party opposite to that of the participant. This could have led pre-existing attitudes to play a role in the observed CIE (Ecker & Rodricks, 2020; Mickelberg et al., 2024). Other studies have found no clear evidence of the CIE on impression formation (e.g. De keersmaecker & Roets, 2017; Ecker & Rodricks, 2020; Mickelberg et al., 2024). Ecker and Rodricks (2020) found that people were capable of correcting their impressions of fictitious characters after allegations of domestic violence were retracted. Similarly, Mickelberg et al. (2024) found that participants fully discounted the retracted information in their vignette even when the behaviour descriptions were congruent or causally related to the misinformation. However, in both these studies, the misinformation was unequivocally retracted leaving no space for participants to evaluate the veracity of the retraction. This fits with O'Rear and Radvansky's (2020) findings, where the CIE only arises if people do not believe in the correction. It is possible that providing a more tentative retraction (MacFarlane et al., 2021) or one from a less reliable source (Ecker & Antonio, 2021; Guillory & Geraci, 2013) could lead to a higher reliance on misinformation. Finally, Cobb et al. (2013), explored the effect of retracting positive misinformation attributed to a political candidate. Here they found that individuals over-corrected their perception of the politician generating a "punishment effect", whereby participants rated the politician more harshly after the correction than in the no-misinformation condition.

The present literature thus shows mixed findings on the CIE in impression formation. In part this might be due to different methodological approaches, where some studies presented misinformation targeting a political candidate (Cobb et al., 2013; Thorson, 2016) while others targeted individuals with no specific partisanship (De keersmaecker & Roets, 2017; Ecker & Rodricks, 2020). Additionally, the certainty with which corrections were presented may have influenced participants' confidence in updating their beliefs, compared to a scenario where the veracity of the correction also needed to be assessed. The latter scenario is more common in people's everyday lives.

To address the mixed findings in this body of literature, our studies systematically vary the reliability of the source making the correction. We explore whether the perceived unreliability of the source providing the retraction might lead to higher rates of reliance on misinformation, as participants may be reluctant to fully update their beliefs when they do not perceive the retraction as credible.

Although there is substantial evidence for the CIE, researchers have yet to determine a consensual account for why this process arises. Providing an explanation for this effect would help limit the detrimental effects of misinformation and provide further insight into developing effective corrections. Corrective methods would not only be advantageous in the political domain but also in the field of science (Lewandowsky et al., 2017; Suarez-Lledo & Alvarez-Galvez, 2021; Swire-Thompson & Lazer, 2020). For example, the spread of misinformation during the COVID-19 pandemic, including false claims about vaccine safety and the origins of the virus, increased vaccine hesitancy and hindered public health efforts (Garett & Young, 2021; Lee et al., 2022). Similarly, misinformation has been shown to undermine climate change initiatives by promoting doubt about the scientific consensus and downplaying the urgency of climate action, making it more difficult to achieve support for necessary policy measures (Chan & Albarracín, 2023; Van der Linden et al., 2017). With regards to economics, this could alleviate the growing concern about the role of misinformation in the disruption of markets (Petratos, 2021) and an increasing concern about the potential costs related to the correction of misinformation and the protection of the public from future exposure (Gradoń, 2020).

1.1. *Explaining the continued influence effect*

The main explanations for the CIE have focused on the role of memory, claiming that the effect emerges from either selective retrieval of the misinformation (Ecker et al., 2015; Ecker et al., 2011; Gordon et al., 2019; Rapp et al., 2014; Rich & Zaragoza, 2016; Swire et al., 2017; also see Ayers & Reder, 1998) or from integration failure when processing the retraction (Brydges et al., 2018; Ecker et al., 2017; Gordon et al., 2017; Kendeou et al., 2014; Kendeou et al., 2019). The former account proposes that the CIE occurs when both accurate and inaccurate information is stored in memory simultaneously. Upon retrieval, misinformation is activated but not effectively suppressed (Ecker et al., 2011). On the other hand, the latter theory suggests that corrections may be poorly integrated into memory for several reasons, such as the difficulty in incorporating new information or the failure to replace the original misinformation with a coherent alternative (Ecker et al., 2022; Susmann & Wegener, 2022). This theory also posits that when the retraction creates a gap in the mental model of an event, individuals may struggle to fill this gap, especially if the correction does not provide a clear alternative explanation for the event's outcome (Connor Desai & Reimers, 2019; Johnson & Seifert, 1994; Rich & Zaragoza, 2016).

An alternative explanation to the two theories outlined above was proposed by Guillory and Geraci (2010), who conducted a study directly asking participants to explain the reason for a retraction. While many participants believed the retraction was a correction of an earlier error, a significant portion suggested that it might be an intentional cover-up. Unsurprisingly, these participants were also more likely to rely on the original information in their inferences. These findings suggest that the CIE might not reflect a cognitive bias, but that underlying reasoning pertinent to the credibility of the retraction might play a part in how

individuals update their beliefs. In this vein, the authors argue that making the correction more believable might reduce participants' reliance on the original misinformation. O'Rear and Radvansky (2020) further developed this hypothesis in their study, where they found that the majority of participants presented with a retraction did not accept it even when the retraction was presented from authority figures. Here they state that "The magnitude of the CIE may be exaggerated not by memory processes from a prior understanding, but from an unwillingness to accept the retraction" (p. 141). This evidence suggests that the underlying processes behind the CIE are more complex than a mere cognitive failure, with factors such as source credibility potentially playing a crucial role.

1.2. The role of source reliability

A range of studies have shown that people are influenced by source reliability when updating their beliefs (Briñol & Petty, 2009; Cone et al., 2019; Kumkale et al., 2010; Pornpitakpan, 2004). Although determining the credibility of a source is complex (Ecker et al., 2022; Lewandowsky et al., 2012), there is a well-documented relationship between source reliability and the acceptance of information. In a meta-analysis spanning five decades, Pornpitakpan (2004) found that a highly credible source is more likely to persuade people than a low-credibility source (e.g. Horai et al., 1974; Hovland & Weiss, 1951; Johnson et al., 1968; Johnson & Izzett, 1969; Lirtzman & Shuv-Ami, 1986; Maddux & Rogers, 1980; Whittaker & Meade, 1968). Classic studies have shown that messages from high-credibility sources are generally more persuasive and more likely to be accepted than those from low-credibility sources (Hovland & Weiss, 1951; Kelman & Hovland, 1953). For example, Hovland and Weiss (1951) found that people tend to accept information from sources they perceive as credible, even when it contradicts their prior beliefs. Individuals are therefore more likely to incorporate new information, even if erroneous, when it comes from a highly credible source (Smith & Ellsworth, 1987; Zhu et al., 2010). Recent studies have further examined how source reliability influences belief revision, especially in the context of corrections. For instance, Walter and Murphy (2018) conducted a meta-analysis of studies that categorised different correction strategies, including those based on source credibility. Their analysis revealed that corrections relying on the perceived reliability of the source tend to have limited effectiveness. They argue that, due to political polarisation and the growing erosion of public trust in official sources, the perceived reliability of a source has little impact on the success of corrections. As Lewandowsky et al. (2012) noted, a source's perceived credibility often depends on pre-existing beliefs: "If you believe a statement, you judge its source to be more credible" (p. 119).

When it comes to the CIE, studies exploring the role of source credibility have yielded inconsistent findings. Dias et al. (2020) found that increasing the visibility of sources on social media and emphasising the reliability of sources—whether participants had previously rated them as reliable or unreliable—does not reduce susceptibility to misinformation. They further conclude that methods aimed at countering misinformation by emphasising source reliability may be ineffective and, in some cases, could even produce counterproductive results. On the other hand, Ecker et al. (2024) found that discrediting a source was partially effective in reducing misinformation reliance. To shed light on this relationship, Walter and Tukachinsky (2020) conducted a meta-analysis exploring how source credibility, among other factors, affects the CIE in the face of correction. Here they find stronger evidence of CIE when the source of misinformation was attributed to a high rather than low credibility source. For the source stating the correction, however, they found no significant difference between high or low levels of source credibility. The evidence they provide therefore suggests that while the credibility of the source of the misinformation affects people's belief in the initial information, the credibility of the source delivering the correction does not.

Another line of research investigating the relationship between source credibility and the CIE was carried out by Guillory and Geraci (2013). Their paper is the most direct exploration of this issue and will be at the core of the present study. Their design used a vignette in which a politician was accused of taking a bribe. To explore whether source credibility plays a role in the evaluation and updating of retracted information, they manipulated the credibility of the retracting sources. In contrast to Walter and Tukachinsky (2020), they found that retracting sources higher in credibility decreased participants' reliance on the original misinformation. They further explored two dimensions of source credibility, expertise and trustworthiness, showing that only the latter affected the persuasiveness of the retraction. These findings were later replicated by Ecker and Antonio (2021). In both sets of studies, however, expertise was defined as "involvement in an event" rather than its more common meaning of "possessing relevant knowledge". To further explore this function of reliability we manipulated sources in terms of trustworthiness and experience rather than involvement.

In addition, both papers showed that participants were less likely to rely on the original misinformation to answer inference questions when the retracting source was highly reliable as compared to unreliable; in both conditions, however, belief in the bribery was not completely extinguished. Here it could be argued that the baseline belief that a politician might take a bribe is higher than zero and therefore it is rational not to completely eradicate belief in bribery. Given two pieces of equally credible information, one claiming the bribe occurred and the other claiming it did not, the correction might therefore be insufficient to bring the belief in the bribery down to a probability of zero, but should bring it back to one's prior belief that bribery has occurred (when no evidence is produced). As also stated by Ecker and Antonio (2021), it would therefore be beneficial in this case to explore the findings using a passive control where there is no claim of a bribery in the story, rather than an active control where the misinformation is stated but not corrected. We therefore suggest that a reexamination of Guillory and Geraci's (2013) measures and controls is crucial to assess the robustness of the CIE. Another aspect of the studies by both Guillory and Geraci (2013) and Ecker and Antonio (2021) is that their designs only manipulated the correcting source, leaving the misinformation source unidentified. This design could drive the difference between the belief in the accusation compared to the belief in the correction, because the original claim might have been perceived as more reliable when expressed as a statement outside of the context of an individual's claim. For example, in one of Ecker and Antonio (2021)'s vignettes the claim reads "Symptoms of inflammatory joint conditions can be treated effectively through remedial yoga" (p. 634) while the retraction states "Debra Phillips, has stated that effective treatments will almost always include pharmaceutical intervention, as practices such as yoga are not effective". In this context it appears that the initial claim is more factual and therefore does not require the identification of a source, while the correction is reported as someone's opinion. This may affect how participants weigh the evidence presented to them. To control for this potential effect, in our study we manipulate both the source of the accuser as well as the corrector, to better represent the reality of such situations.

1.3. Rational frameworks for the CIE

Recent analyses of the CIE acknowledge that failing to revise one's beliefs in light of new evidence is not always illogical and thus that the CIE might sometimes be rational (e.g. Connor Desai et al., 2020; Gershman, 2019; Haselton et al., 2009; Pilgrim et al., 2024). Connor Desai et al. (2020) argue that the characteristics of the source of information are essential to the evaluation of content: if a source is likely to lie or make errors, it may be completely reasonable to discard their claims. Therefore, when the source of misinformation is perceived as more credible than the source of the correction it might be rational not to update one's beliefs (Connor Desai et al., 2020; also see Jern et al., 2014). In this vein, empirical studies have shown that people do in fact incorporate their personal assessment of a source's credibility when

evaluating testimony (Harris et al., 2016; Harris & Hahn, 2009; Madsen, 2016; Merdes et al., 2021) and sensibly alter a source's reliability if new information is made available (Madsen et al., 2020). This process is not only prudent but suggests rational updating. To test this proposal Connor Desai et al. (2020) apply a Bayesian network formalism, conceptualising the CIE using a scenario with contradictory testimonies. They include crucial components such as the perceived reliability of sources, both misinforming and correcting, as well as the concept of temporal dependence (i.e. the assumption that misinformation must precede its retraction). In testing this framework, they show that on average individuals intuitively make assumptions that would categorise the CIE as a rational process. Participants also appropriately penalised the reliability of a source when it was contradicted. This study links to other research on how individuals evaluate conflicting evidence (e.g. Fenton et al., 2013), where Bayesian networks (BNs) were shown to be useful tools for modelling legal arguments. Other papers have also conceived of the issue of misinformation correction as a Bayesian inference problem. Zmigrod et al. (2023), for example, propose a Misinformation Receptivity Framework (MRF) aimed at formalising the interactions between the cognitive and communicative mechanisms that govern susceptibility to ideological misinformation. They suggest that misinformation receptivity is a Bayesian inference problem which is modulated and distorted by ambiguous or deceptive information. In line with this formalisation, models of source reliability, such as those by Bovens and Hartmann (2003) and Olsson (2013) also explore rational mechanisms in the context of belief updating (Merdes et al., 2021). These models offer a systematic Bayesian approach to evaluating information by considering factors such as the independence and reliability of sources, as well as the plausibility and coherence of the information itself. The main argument in Bovens and Hartmann (2003) is that individuals are rationally required to update both their beliefs and their trust in a source based on a given report, especially when the source's reliability has not been externally established (Merdes et al., 2021; also see Jarvstad & Hahn, 2011). These models have been applied to a range of issues including evidence, testimony and voting.

Following the approach of these papers, we maintain that exploring the rational aspect of CIE in further depth could improve our understanding of belief updating and better inform initiatives for misinformation correction. While previous models of the CIE have touched on the rationality of belief persistence, there has been limited exploration of the role of source reliability in belief updating. Our study addresses this gap by examining the complexities of belief updating, particularly in relation to the perceived credibility of sources. Our paper clarifies conflicting evidence on the role of source reliability and provides new insights into how rational belief updating can occur in the context of misinformation correction.

1.4. The present study

The present study aims to shed light on the mixed evidence for CIE. The first two experiments will build on the findings from Guillory and Geraci (2013), re-examining their measures and controls. In parallel with their original study, we will aim to replicate their hypotheses with one key change. In accordance with Connor Desai et al. (2020) we anticipate that participants receiving a correction from a highly credible source will effectively reduce their reliance on misinformation back to baseline. The subsequent experiments will manipulate the source reliability of both the misinformer and the corrector to test persuasion efficacy, in particular examining whether participants update their beliefs rationally. Here we anticipate that both the source alleging the bribery and the source providing the correction will influence the extent to which people rely on the belief that the bribery occurred. In line with prior literature, we also predict that individuals will update their subjective perception of the reliability of the sources after contradiction. Finally, contrary to Guillory and Geraci (2013) we expect that expertise as well as trustworthiness will affect whether people update their beliefs in the misinformation.

2. Experiment 1

The first experiment used Guillory and Geraci's (2013) study design in the attempt to replicate their findings. Participants read a vignette presented as a series of messages about a politician running for re-election. Half-way through the vignette participants read that a report claimed that the politician was seen taking bribe money. (As there is no ground truth on whether this claim is in fact misinformation, we will refer to it as an allegation throughout the experiment related sections). In the control condition, this allegation was not corrected. In the correction condition, at the end of the story, participants read a statement indicating that the allegation was shown to be untrue (see Appendix A). Compared to the original Guillory and Geraci (2013) study we added a third condition where there was no mention of the bribe. This was added to identify the baseline probability of the politician taking a bribe when there was no accusation and consequently no correction. After reading the story, participants rated the likelihood of the politician having taken the bribe, as well as the reliability of the source of the report. Our aim was to replicate the findings of Guillory and Geraci (2013) examining the effect of a "correction" to the story, negating the bribe. As no source was identified in this context, this study also sheds light on the default reliability attributed by participants to the originator of the claim. It further provides a base for comparison for the subsequent studies in which the sources involved are partially (Experiment 2) or fully identified (Experiment 3 and 4).

We also conducted a norming study, after the bribe likelihood and reliability questions, to identify which sources participants considered to be the most reliable within the particular context of the story. We analysed two elements of credibility: trustworthiness and expertise.

2.1. Methods

2.1.1. Participants

101 US based participants were recruited on Prolific (N ≈ 33 per condition), and randomly assigned to one of the conditions. The sample size parallels that of Guillory and Geraci (2013). The age range was between 18 and 76 and evenly distributed between those who identified as female and male. Participants were paid £1.50 (~$1.97) for their time (Median time to complete = 8:01 min).

2.1.2. Design

The study used a between-subjects design with condition (Control, Correction, No Allegation) as the manipulated variable. The dependent variables were the likelihood that the bribery took place and the reliability rating of the source making the bribery allegation (see https://osf.io/rhjnx).

2.1.3. Materials and procedure

Participants were given a story to read about a politician seeking re-election (see Appendix 1). In two situations (Control and Correction), there was a significant detail introduced – that the politician was observed accepting bribes. This information came from an unnamed source. In the No Allegation scenario where there was no bribery allegation, the message was replaced with a description of the politician having a debate with his opponent. Of the vignettes mentioning the bribe, the Correction condition included a final message explaining that the previous report was incorrect, and the politician didn't receive any bribes. The Control and No Allegation versions didn't include such a correction. Instead, the twelfth message talked about local school children's interest in following the election coverage.

Participants were informed at the start of the study that they would be asked to read and recall a story. They were shown one message at a time, and they could read at their own pace, but they couldn't go back to previous messages. After finishing the whole story, to test their attention

they were asked three randomised multiple-choice questions about what happened in the story. Only participants who got at least two out of three questions correct were included in the analysis. This criterion was determined a priori to ensure a minimum level of engagement with the task. A total of 5 participants were excluded based on this criterion, resulting in a final sample size of 101. The exclusion of these participants did not significantly alter the outcome of the analysis, as the overall pattern of results remained consistent. Subsequently, participants were asked to rate, on a scale from 0 to 100, how likely it was that the politician accepted a bribe during the election campaign and how reliable the source was that reported the bribery. The decision to use a scale from 0 to 100, diverging from the Likert scale of 1 to 7 in Guillory and Geraci's (2013) study was made for a number of reasons. Firstly, this broader range offers finer distinctions in participants' judgments, enhancing precision and sensitivity. It also facilitates straightforward analyses and comparisons across studies, reducing potential ceiling or floor effects and improving variability for statistical robustness and generalizability. Overall, the 0 to 100 scale aligns with our aim of capturing comprehensive assessments, enhancing the interpretability and reliability of our results compared to the narrower Likert scale. This rationale draws upon insights from prior research utilising similar scales (e.g., Sung & Wu, 2018, Dourado et al., 2021), highlighting the advantages of this approach in capturing nuanced responses and facilitating robust statistical analyses.

In the norming section participants were then given a list of 28 sources, paralleling those from the original study, and asked to imagine that the correction in the story came from the source in question. (Participants in the control conditions were told prior to the rating that the correction had occurred.) They were then asked to rate the expertise and trustworthiness of each source on a scale of 0–10. Following Guillory and Geraci (2013), trustworthiness was defined as the willingness of a source to provide accurate and reliable information, whereas expertise was defined as the extent to which a speaker is capable of making correct assertions. The order of the sources was randomised as well as whether participants were asked to rate the expertise or the trustworthiness of the sources first.

2.2. Results and discussion

2.2.1. Examining allegation correction

A one-way analysis of variance (ANOVA) was conducted to assess the differences in the likelihood ratings among the three conditions: Control, Correction, and No Allegation. Post-hoc comparisons were performed using Tukey's Honestly Significant Difference (HSD) test to determine specific pairwise differences between conditions. Significance for all analyses was set at $p < 0.05$.

The ANOVA revealed a significant main effect of the Vignette condition on the likelihood ratings of the politician taking a bribe ($F(2, 98) = 12.34$, $p < 0.001$). Post-hoc comparisons using Tukey's HSD test were conducted to further explore the differences between the three Vignette conditions: No Allegation, Control, and Correction (see Fig. 1).

Allegation vs. No Allegation: Participants who read about the bribery allegation with no correction rated the likelihood of the politician taking a bribe significantly higher compared to those who did not read about the allegation (Mean Difference = 19.11, 95 % CI [4.32, 33.91], $p = 0.008$).

No Allegation vs. Correction: There was no significant difference in the likelihood ratings between the No Allegation and Correction conditions (Mean Difference = 0.29, 95 % CI [−14.50, 15.09], $p = 0.999$). This suggests individuals' perception of the bribery after correction of the allegation is comparable to those who never read the allegation in the first place.

Control vs. Correction: Finally, participants in the Correction condition rated the likelihood of the politician taking a bribe significantly lower compared to those who read about the allegation with no correction (Mean Difference = −18.82, 95 % CI [−33.72, −3.91], $p = 0.009$).

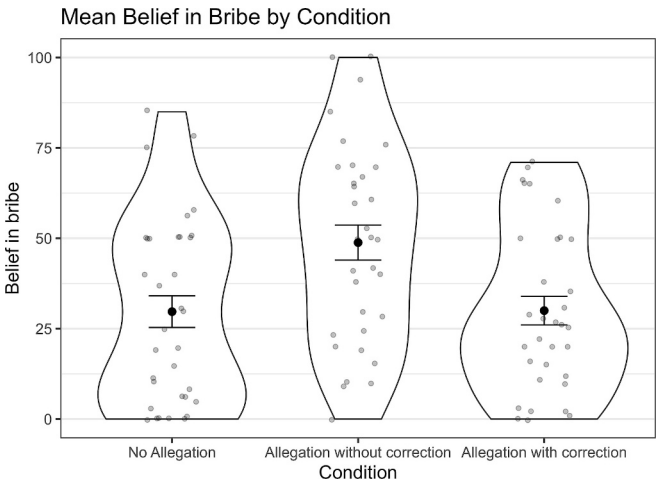


Fig. 1. Mean belief in bribe by condition (No Allegation, Allegation without correction, Allegation with correction) for Experiment 1. Note. Error bars represent 95 % confidence intervals. Ratings ranged from 0 (very unlikely) to 100 (very likely). Individual data points have been jittered along the x-axis to improve clarity and visibility.

These findings suggest that the presence of an explicit correction significantly reduced the perceived likelihood of the politician taking a bribe compared to scenarios with no correction. Conversely, participants in the control condition, where no correction was provided, rated the likelihood of bribery significantly higher than those in the correction condition. Notably, there was no significant difference between the condition where the allegation was absent and the condition where it was corrected, suggesting that the correction effectively restored beliefs to baseline levels. These findings replicate Guillory and Geraci's (2013) results, which indicate that, even after a bribery allegation is discredited, the belief that bribery could have occurred does not fully return to zero. However, by introducing a condition where no allegation was made, it becomes evident that individuals may indeed be discounting the discredited information and reverting to baseline belief level. Thus, while Guillory and Geraci's findings suggest that retracting the bribery claim does not eliminate suspicion entirely, our analysis highlights that rational participants may not necessarily expect the likelihood to drop to zero, as baseline expectations of bribery—without any specific allegation—tend to remain noticeably above zero. It should also be noted that the allegation itself had quite a modest effect, with the average belief in the bribe after the claim was made averaging around 50 %. This surprising effect can be attributed to two factors. Firstly, the politician in the vignette appeared to be quite likeable, as discussed in the qualitative analysis section (see section 6) which could have affected people's willingness to accept negative evidence against him. Secondly, the claim, replicated from the original study, was made from an unknown source. This could have further increased scepticism in the claim. This possibility underscores the importance of examining the effects of varying both the source of the accusation and the source of the retraction, as explored in the following sections.

Additionally, a linear regression analysis revealed a significant relationship between the perceived likelihood of the politician's involvement in bribery and the perceived reliability of the source making the allegation ($\beta = 0.80$, $p < 0.001$).The overall model fit was significant, with approximately 47.82 % of the variance in the perceived likelihood of bribery being explained by the perceived reliability of the source. The findings suggest a strong positive association between the perceived reliability of the source making the bribery allegation and the perceived likelihood of the politician's involvement in bribery, supporting the hypothesis that perceived credibility of the source affects the extent to which individuals believe in the claim.

2.2.2. Norming

A descriptive analysis was conducted to evaluate participants' rating of trustworthiness and expertise of 28 sources (see Fig. 2). Three sources that on average rated high on both expertise and trustworthiness and three sources that rated low on both were selected for Experiment 2 (see Table 1). For the reliable sources these were a government report, a representative of the legal state department and a district attorney. For the less reliable sources these were a celebrity, a political satire news channel and the wife of the politician's opponent.

3. Experiment 2

In the second study, we aimed to replicate Guillory and Geraci's (2013) experiment, investigating whether individuals could correct their erroneous beliefs when the correction was delivered by a reliable source. Participants were presented with the same narrative as Experiment 1 portraying a politician seeking re-election through a series of messages (see Appendix A). One message conveyed that the politician was observed accepting bribes. In the control group, this allegation remained uncorrected, while in two correction groups, it was corrected. In one correction group, the correction (that the politician did not accept bribes) was provided by one of three highly credible sources identified through the norming in Experiment 1 (which assessed trustworthiness and expertise). These were, as in the original study, a government report, the district attorney, or a representative from the state legal department. In the alternative correction group, the correction was provided by one of three less credible sources identified using the same assessment criteria: a political satire news channel, a celebrity actor, or the wife of the politician's opponent. In the original study, these sources were identified as a popular political blogger, a celebrity actor, or an interest group supporting the politician. Here, the variance in the rating of less credible sources between the original study and our current research may be attributed to a decade-long gap in which source perceptions might have altered. The primary focus of this study was to determine whether participants would be significantly less inclined to retain the original allegation if the correction originated from a highly credible source compared to a less credible source or no correction at all. Secondly, we wanted to test whether in all conditions there was evidence of the Continued Influence Effect for which belief in the allegation would be above baseline even when a credible correction was delivered. Our hypotheses were therefore the following:

Hypothesis 1. Higher levels of reported trustworthiness and expertise of the correcting source would decrease belief in the allegation.

Hypothesis 2. Following correction, belief in the allegation would be significantly above baseline.

Hypothesis 3. The likelihood of voting for the political candidate would correlate with the belief in the correction.

3.1. Methods

3.1.1. Participants

100 US based participants were recruited on Prolific ($N = 33$ per condition), and randomly assigned to one of the conditions. The sample size parallels that of Guillory and Geraci (2013). The age range was between 20 and 77 with a mean of 41.97 (SD = 15.86) and evenly distributed between those who identified as female and male. Participants were paid £2.50 (~$3.15) for their time (Median time to complete = 18.02 min).

3.1.2. Design

The study used a between-subjects design with condition (High reliability, Low reliability and Control) as the manipulated variable. The dependent variables were the participants' estimation of the likelihood that bribery took place and the reliability rating of the source making the bribery allegation (see https://osf.io/rhjnx).

3.1.3. Materials and procedure

As with the first experiment, all participants were given a story to read about a politician seeking re-election (see Appendix A). Each version contained a crucial detail: that the politician was accused of accepting bribes, sourced from an unnamed informant. In two versions (the correction conditions), participants received a final message stating that the accusation was incorrect, and the politician hadn't accepted any bribes. In these correction conditions, participants received the correction from one of six named sources, with three sources in each of the high and low credibility groups. The specific source was randomly assigned to each participant (see Appendix B). In the high credibility correction condition, the correction came from a source rated high in both expertise and trustworthiness. In the low credibility version, the correction came from a source rated low in both expertise and trustworthiness. The control version didn't include a correction message; instead, the final message talked about local school children following the election coverage. The survey procedure was identical to Experiment 1. Once they had read all the survey messages, participants rated the likelihood of the politician accepting a bribe, as well as the reliability of both the source making the allegation and the source making the correction. Subsequently, they rated their likelihood of voting for the politician and provided reasons for their voting decision. They were also asked to explain why they believed there was a correction in the story, following the method used in previous studies (Guillory & Geraci, 2010; Guillory & Geraci, 2013). Finally, all participants stated their political orientation.

3.2. Results and discussion

3.2.1. Examining information correction based on source reliability

A one-way analysis of variance (ANOVA) was conducted to assess the differences in likelihood ratings among three conditions: High reliability, Low reliability, and Control (see Fig. 3). Post-hoc comparisons were performed using Tukey's Honestly Significant Difference (HSD) test to determine specific pairwise differences between conditions. For all analyses, significance was set at $p < 0.05$.

The mean likelihood rating for participants exposed to corrections from more reliable sources was $M = 29.2$, $SE = 5.13$, 95 % CI [19.0, 39.4]. Conversely, those receiving corrections from less reliable sources reported a significantly higher mean likelihood rating of $M = 50.5$, $SE = 5.21$, 95 % CI [40.2, 60.9], indicating a substantial impact of source credibility. Participants in the control condition, without any correction, exhibited a mean likelihood rating of $M = 53.5$, $SE = 4.83$, 95 % CI [43.9, 63.1], slightly higher than those in the unreliable condition. Tukey's HSD test revealed significant differences in likelihood ratings between various conditions:

Low vs. High reliability: Participants in the Low reliability condition rated the likelihood of the politician accepting bribes significantly higher than those in the High reliability condition (Mean Difference = 21.36, 95 % CI [3.97, 38.76], $p = 0.012$), indicating a perceived higher impact of corrections from more reliable sources.

Control vs. High Reliability: Likelihood ratings in the Control condition were significantly higher compared to the High reliability condition (Mean Difference = 24.34, 95 % CI [7.57, 41.11], $p = 0.002$), suggesting the influential role of correction credibility.

Control vs. Low reliability: There was no significant difference in likelihood ratings between the Control and Low reliability conditions (Mean Difference = 2.98, 95 % CI [−13.93, 19.89], $p = 0.908$), indicating similar perceptions of likelihood when corrections came from less reliable sources compared to no correction.

These results underscore the significant impact of correction source credibility on participants' perceptions of a politician's likelihood to accept bribes. In parallel with the first hypothesis, corrections from less reliable sources led to significantly higher likelihood ratings compared

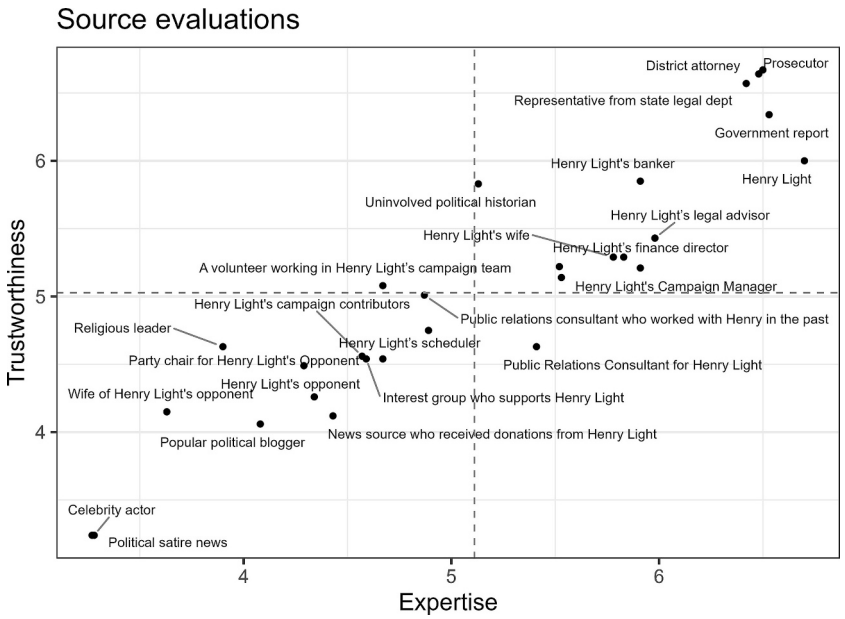


Fig. 2. Average source evaluations rated for expertise and trustworthiness.

Table 1
Mean trustworthiness and expertise of sources correcting the claim of bribery.

Source	Mean Expertise	SD	Mean Trustworthiness	SD
Celebrity	3.28	2.76	3.24	2.52
Political Satire news	3.27	2.85	3.24	2.63
Wife of politician's opponent	3.63	2.98	4.15	3.17
Government report	6.53	2.33	6.34	2.49
District attorney	6.48	2.46	6.64	2.30
Representative from state legal department	6.42	2.43	6.57	2.34

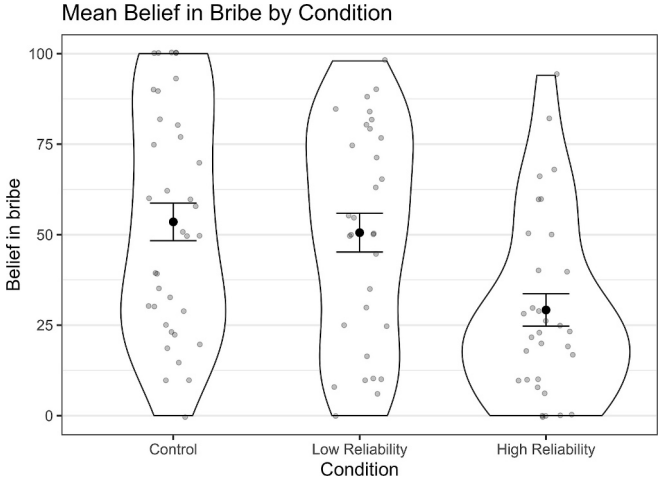


Fig. 3. Mean belief in bribery ratings across experimental conditions (Control, Low reliability, and High reliability). Note. The horizontal axis indicates the reliability manipulation of the corrector, while the vertical axis displays the average belief in bribery, with error bars representing 95 % confidence intervals. Ratings ranged from 0 (very unlikely) to 100 (very likely). Individual data points have been jittered along the x-axis to improve clarity and visibility.

to corrections from more reliable sources, emphasising the importance of source credibility in correcting misinformation.

Additionally, to assess whether the correction could bring the likelihood of the politician having taken the bribe back to baseline, we compared ratings of bribery likelihood in the High reliability condition to those in the no-allegation condition from Experiment 1.

High reliability vs. Control-No Allegation: The mean likelihood rating for participants in the High reliability condition was $M = 29.2$, $SE = 4.81$, 95 % CI [19.7, 38.7]. Similarly, participants in the no-allegation condition reported a similar mean likelihood rating of $M = 29.7$, $SE = 4.67$, 95 % CI [20.5, 38.9]. A Tukey multiple comparisons of means test revealed that there was no significant difference in likelihood ratings between the High Reliability and No Allegation conditions (Mean Difference = 0.52, 95 % CI [−17.95, 18.99], $p = 0.999$), suggesting that the correction from a highly reliable source effectively restored beliefs to baseline levels.[1]

These findings indicate that the correction from a highly reliable source was successful in mitigating the impact of the bribery allegation, bringing the likelihood of the politician having taken the bribe back to baseline. In contrast with the original study we therefore did not find that the Continued Influence Effect is ubiquitous to all conditions, but that participants are capable of reducing the likelihood of the bribery back to baseline when given a highly reliable correcting source.

3.2.2. Translation into voting intention

To evaluate the hypothesis regarding the association between the likelihood of voting for the political candidate and belief in the correction, a linear regression analysis was performed. The results of the regression model indicated a significant positive relationship between the likelihood of voting for the political candidate and belief in the correction (Reliability rating of the correcting source) ($\beta = 0.44$, $p < 0.001$). This implies that as participants' belief in the correction increased, their likelihood of voting for the candidate also tended to increase. The overall model fit was significant, with the predictor

[1] While we acknowledge that having a no-allegation condition in Experiment 2 could have provided additional clarity, we believe that the methodological consistency and similar participant pools provide a solid foundation for making valid comparisons across the experiments. Therefore, we maintain that the comparisons made in the manuscript are both appropriate and meaningful

variable (Reliability rating of the correcting source) explaining approximately 19.24 % of the variance in the likelihood of voting for the political candidate (*Multiple R-squared* = 0.19, *Adjusted R-squared* = 0.18). These results align with those of the original study, indicating a moderate positive correlation between participants' belief in the correction and their likelihood of voting for the political candidate. Further analysis on voting measures by condition can be found in the appendix (see Appendix C).

4. Experiment 3

In the third experiment, we departed from Guillory and Geraci's (2013) original study and explored a more complex scenario examining how both the source alleging bribery and the source denying it interact to influence belief in the occurrence of bribery. Unlike the previous study where the source making the claim remained unidentified, here we identified both the accuser and the corrector. We manipulated the trustworthiness and experience of both sources to observe how individuals adjusted their beliefs in response to different combinations of sources. The choice to manipulate experience as a function of expertise is based on prior evidence suggesting experience is an effective indicator of perceived competence. Research suggests that individuals tend to infer expertise based on observable cues such as the length and diversity of an individual's experience in a particular domain. These cues can influence perceptions of competence, leading individuals to view those with more extensive experience as more knowledgeable and skilled in their field (Ames & Kammrath, 2004; Bandura, 1977; Fiske & Neuberg, 1990). These findings suggest that experience can indeed be an effective measure of perceived expertise, influencing how individuals evaluate the knowledge and skill of others.

With regards to the study design, the vignette remained consistent with the previous study, except that both the bribery accusation and the correction originated from a prosecutor with varying levels of trustworthiness and experience (see Table 2). We chose to use prosecutors based on their high average ratings of trustworthiness and expertise in the norming study (Mean expertise = 6.50, SD = 2.44; Mean trustworthiness = 6.67, SD = 2.37). We manipulated expertise by varying how much experience the prosecutor had in investigating corruption, either having no experience or extensive experience. Trustworthiness was manipulated by stating that the prosecutor either had a clean record or previous allegations of wrongdoing (see Table 2).

In addition to the rating questions from the previous study, participants were also asked to rate the reliability of the first source and belief in the probability of the bribe after the allegation. This resulted in two sets of ratings: one for the perceived reliability of the first source and one for the likelihood of the bribe, both before and after the correction. The aim here was to gain a better understanding of how participants update their beliefs after they receive the correction.

There were eight experimental conditions, as well as a control condition where the vignette remained unchanged, but with no information about the trustworthiness or experience of either prosecutor. The eight conditions involved the full factorial of all possible combinations of reliability between the first and second sources in terms of either experience or trustworthiness. If the claim was made by a reliable prosecutor, the correction could be delivered by either a reliable or unreliable prosecutor. Conversely, if the claim was made by an unreliable prosecutor, the correction could be made by either a reliable or unreliable prosecutor. Among the eight conditions, reliability was based on experience in four of them, while in the other four, it was based on trustworthiness.

Table 2
Manipulating experience and trustworthiness of a prosecutor.

Information	Low	High
Experience	No experience investigating corruption	Extensive experience investigating corruption
Trustworthiness	Allegations of previous wrongdoings against him	Clean record

Experiment 2 indicated that individuals could correct erroneous information if the correction came from a credible source. However, since the source making the allegation was unidentified, there was no comparison or assessment between the two sources. In this case, source reliability influenced their evaluation of the correction but not the original allegation, as the latter lacked an identified source. The current experiment aims to investigate whether a correction from a reliable source influences belief in the allegation independently of the source making the allegation. We hypothesise that allegations from less trustworthy or experienced sources would be more susceptible to correction if the correction comes from a trustworthy or experienced source. Conversely, there should be a stronger adherence to the original information if the accuser is reliable while the corrector is not.

Hypothesis 1. The reliability of both the source alleging the bribery and the source providing the correction would influence the extent to which individuals rely on the belief that the bribery occurred.

Hypothesis 1a. If the allegation is made by a reliable source, the belief that the bribery took place would be higher than if it is made by a less reliable source.

Hypothesis 1b. If the source making the bribery claim remains constant, participants receiving the correction from a more reliable source would show reduced reliance on the original information compared to those receiving the correction from a less reliable source.

While Guillory and Geraci (2013) suggest that trustworthiness plays a central role in influencing belief in allegations, we propose that expertise, manipulated through the experience of the source, could also significantly affect belief revision. However, in line with previous findings (e.g. Ecker & Antonio, 2021), we argue that trustworthiness will exert a greater influence on the extent to which individuals revise their beliefs and accept corrective information.

Hypothesis 2. While experience will contribute to belief revision, trustworthiness will have a greater impact on the extent to which individuals revise their beliefs and accept corrective information.

Finally, we anticipate that the contradiction provided by the correcting source would lead to a decrease in the reliability rating of the source making the initial allegation. This hypothesis draws from research by Connor Desai et al. (2020) on the rational Continued Influence Effect, which demonstrates that following a rational process, lay reasoners appropriately penalise the reliability of sources that contradict the initial information.

Hypothesis 3. The reliability rating of the source making the allegation would decrease following contradiction by the correction.

4.1. Methods

4.1.1. Participants

597 US based participants were recruited on Prolific (N = 60–90 per

condition[2]), and randomly assigned to one of the 9 conditions. The sample size was doubled from the previous study to account for the increase in the number of factors we are analysing. Given that manipulating the reliability of the sources in terms of trustworthiness or expertise did not result in a statistically significant difference, the analysis was conducted on five conditions ($N = 108$ per condition). In this revised analysis, we combined the high trustworthiness and high expertise conditions into a single 'high reliability' group, and the low trustworthiness and low expertise conditions into a 'low reliability' group. This resulted in a full factorial design with four conditions (low-low, low-high, high-low, and high-high) and one control condition. The age range was between 20 and 82 with a mean of 42.09 (SD = 13.56) and evenly distributed between those who identified as female and male. Participants were paid £0.45 (~$0.57) for their time (Median time to complete = 3.26 min). Overall, 21 respondents were removed as they did not abide by the pre-established criterion of minimum 2 correct multiple choice questions out of 3.

4.1.2. Design

The study adopted a between-subjects design featuring eight conditions as the manipulated variable (see Table 3a and 3b). The dependent variables were the likelihood that the bribery took place and the reliability rating of the source making the bribery allegation as well as that of the source delivering the correction (see Appendix D).

4.1.3. Materials and procedure

Participants were presented with a vignette portraying a politician seeking re-election, resembling the one in the previous experiments. However, in this experiment, the source alleging bribery was identified as a prosecutor (with levels of experience or trustworthiness varied between conditions). Conversely, the correction at the end of the vignette, stating that the politician did not accept the bribe, was delivered by a different prosecutor (also with varying levels of experience or trustworthiness according to condition). A control condition was also included, where the experience and trustworthiness of both prosecutors were not given. Following the allegation, participants rated the reliability of the source on a scale of 0–100, ranging from very unreliable (0) to very reliable (100). Subsequently, participants rated the likelihood of the politician accepting the bribe on a 0–100 probability scale. After receiving the correction, participants were again asked to rate the likelihood of the politician accepting the bribe, along with their ratings of the reliability of both the source alleging the bribe and the source providing the correction. An open text box was provided for participants to explain the reasons behind their belief ratings regarding the bribery claim. The qualitative data that was collected is discussed in section 6. Lastly, participants were asked to indicate their political orientation.

Table 3a
Conditions manipulating experience of prosecutors.

Condition	Prosecutor making the allegation	Prosecutor making the correction
1	No experience	No experience
2	Extensive experience	Extensive experience
3	Extensive experience	No experience
4	No experience	Extensive experience

Table 3b
Conditions manipulating trustworthiness of prosecutors.

Condition	Prosecutor making the allegation	Prosecutor making the correction
1	Low trustworthiness	Low trustworthiness
2	High trustworthiness	High trustworthiness
3	High trustworthiness	Low trustworthiness
4	Low trustworthiness	High trustworthiness

4.2. Results and discussion

4.2.1. No difference between trustworthiness and experience

A repeated-measures ANOVA examined the influence of trustworthiness and experience on participants' belief in bribery. This analysis revealed no significant differences in the effect of either type of reliability information on participants' belief that the bribery occurred before ($F(1, 476) = 0.112$, $p = 0.738$) or after correction ($F(1, 548) = 0.031$, $p = 0.859$). These results suggest that whether source reliability was a function of trustworthiness or experience did not make a difference. Further analysis of the trustworthiness and experience manipulations can be found in the appendix (see Appendix E). Consequently, the subsequent analysis will explore overall reliability, irrespective of whether it was based on trustworthiness or experience.

4.3. Judgments before correction

4.3.1. Both experience and trustworthiness affect ratings of source reliability of the source making the allegation

A repeated-measures ANOVA indicated significant effects of both prosecutor experience and trustworthiness manipulations on participants' perceived reliability of the source making the allegation ($F(1, 476) = 66.234$, $p < 0.001$). Participants rated the source as more reliable when it had high experience and trustworthiness ($M = 58.3$, $SE = 1.48$) compared to low experience or trustworthiness ($M = 41.2$, $SE = 1.49$), with a mean difference of 17.1 ($SE = 2.11$, $p < 0.001$). (See Fig. 4).

4.3.2. Reliability of the source making the allegation affected belief in bribery

A repeated-measures ANOVA revealed a significant effect of the reliability of the source making the allegation on participants' likelihood

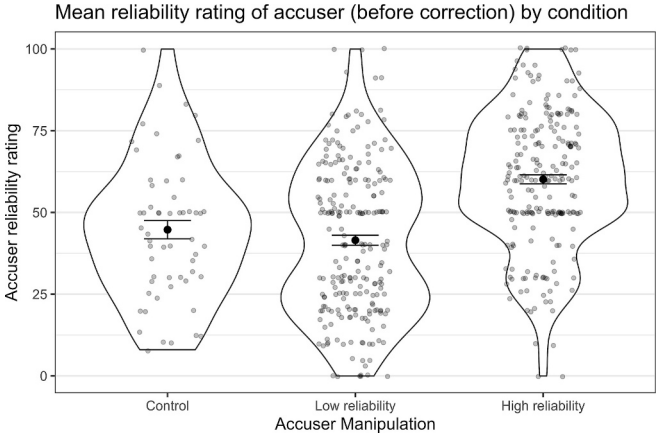


Fig. 4. Mean reliability ratings of the accuser (before correction), categorised by accuser reliability conditions. Note. Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.

[2] In the initial implementation of Experiment 3, unequal sample sizes were introduced as a result of the experiment missing two key conditions in the factorial design. After recognizing the issue, the experiment was rerun with 8 conditions and increased power. Despite the unequal sample sizes between the two runs of the experiment, random assignment was maintained throughout, ensuring unbiased allocation of participants across conditions. To ensure that the unequal sample sizes did not impact the results, an additional analysis was conducted using matched sample sizes for each condition. The results from this analysis were consistent with the original findings and confirmed that the pattern of significant and non-significant results remained unchanged. The supplementary analysis will be made available on OSF for transparency and further examination.

ratings of the bribery occurrence before the correction (F(1, 476) = 33.578, $p < 0.001$). Participants rated the likelihood of the bribe being taken significantly higher when the source had high reliability compared to low reliability (mean difference = 11.75, $p < 0.001$). (See Fig. 5).

4.3.3. Participants' perceived source reliability influenced belief in bribery allegation

A linear regression analysis showed a significant positive association between participants' ratings of the reliability of the source making the bribery allegation and their belief in the bribery occurrence before receiving the correction ($\beta = 0.642$, $SE = 0.0311$, $t = 20.623$, $p < 0.001$). The regression model was highly significant (F(1, 476) = 425.3, $p < 0.001$), explaining a substantial amount of variance in participants' beliefs ($R^2 = 0.472$).

4.4. Judgments at second stage (after correction)

4.4.1. Experience and trustworthiness affect source reliability ratings of corrector

An ANOVA revealed a significant effect of the presented reliability of the source making the correction on participants' reliability ratings of that source (F(1, 548) = 43.354, $p < 0.001$). Participants rated the source significantly higher in reliability when information identified them as trustworthy or experienced ($M = 64.4$, $SE = 1.38$) compared to when they were identified as untrustworthy or inexperienced ($M = 51.5$, $SE = 1.38$), with a mean difference of 12.8 ($SE = 1.95$, $p < 0.001$). (See Fig. 6).

4.4.2. All sources were penalised by contradiction

Paired sample *t*-tests indicated statistically significant decreases in reliability ratings of the source making the allegation after receiving the correction across all conditions ($p < 0.001$) (See Fig. 7). The magnitude of the drop did not significantly differ across conditions ($p > 0.05$).

4.4.3. Accuser reliability affects belief in bribery - corrector reliability does not

An ANOVA was performed to examine the impact of first and second source reliability, along with timepoint (pre- and post-correction), on participants' beliefs regarding bribery occurrence (P(Bribe)). Results showed a significant main effect of the reliability information of the accuser (F(1, 429) = 32.92, $p < 0.001$), indicating its influence on participants' beliefs. However, there was no main effect of the reliability of the corrector (F(1, 429) = 0.85, $p = 0.358$), suggesting no significant

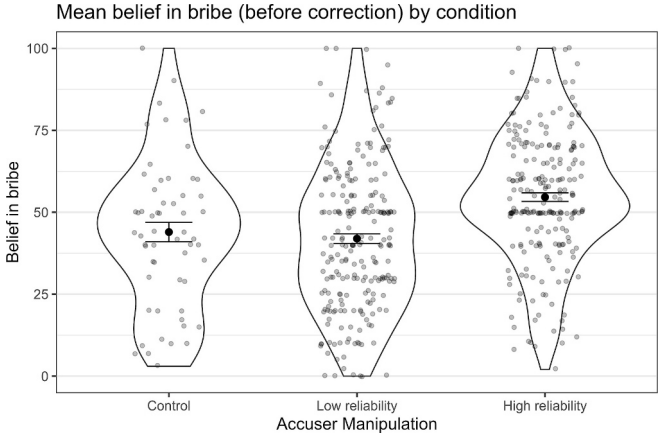


Fig. 5. Mean belief in bribery (before correction), categorised by accuser reliability conditions. Note. Error bars represent the standard error of the mean (SE). Jittered points illustrate individual ratings, indicating variability within conditions.

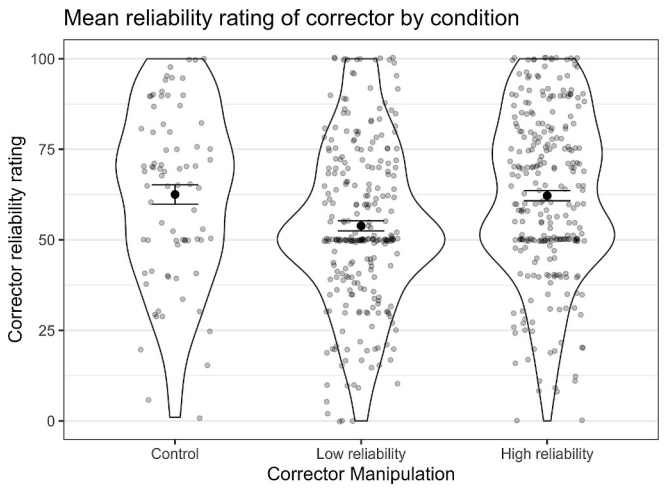


Fig. 6. Mean reliability ratings of the corrector, categorised by accuser reliability conditions. Note. Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.

effect on beliefs about bribery occurrence, and there was no interaction between the two reliability factors ($p > 0.05$), indicating independent effects of these factors on beliefs. Additionally, a main effect of timepoint was observed (F(1, 429) = 225.56, $p < 0.001$), reflecting significant changes in beliefs from allegation to correction receipt. Finally, a significant interaction effect between accuser's reliability information, corrector's reliability information, and timepoint was found (F(1, 429) = 6.88, $p = 0.009$), but subsequent analyses revealed no significant differences in the magnitude of belief change (smallest $p = 0.09$), suggesting a potentially negligible overall impact of this interaction.

These findings suggest that although there is a relationship between the reliability of the second source and the belief in the bribery occurring after correction, the manipulation of reliability might not be strong enough to capture this relationship in full.

4.4.4. Reduction in belief in bribery post-correction is consistent across conditions, irrespective of source reliability

Belief in the bribery occurrence decreased after correction in all conditions. We used pairwise comparisons to assess the difference in these reductions (See Fig. 8). Across all conditions, the reduction in belief about bribery from before to after the correction was not statistically significant (estimate = −1.33, $SE = 1.92$, df = 429, t-ratio = −0.695, $p = 0.487$), indicating that while all conditions experienced a decrease in belief, the size of the reduction did not vary significantly based on the reliability of the correcting source. This suggests that while receiving correction affected beliefs about bribery, the reliability of the correcting source did not. Once again, these results hint at the need for a more effective manipulation to understand the underlying dynamics.

5. Experiment 4

In this final experiment, we aimed to deepen our understanding of how people updated their beliefs by manipulating source reliability more robustly. In Experiment 3, we manipulated the reliability of the source alleging the bribery and the source making the correction by considering either their expertise or trustworthiness. In contrast, in the current experiment, we combined both factors: the reliability of each prosecutor was therefore based both on (i) their experience with corruption investigations, and (ii) whether they had a clean record or were previously alleged to have committed wrongdoings (see Appendix F). The assumption in combining these qualities was not necessarily that there would be an additive effect on reliability but that source reliability would become more salient as a relevant factor in evaluating the

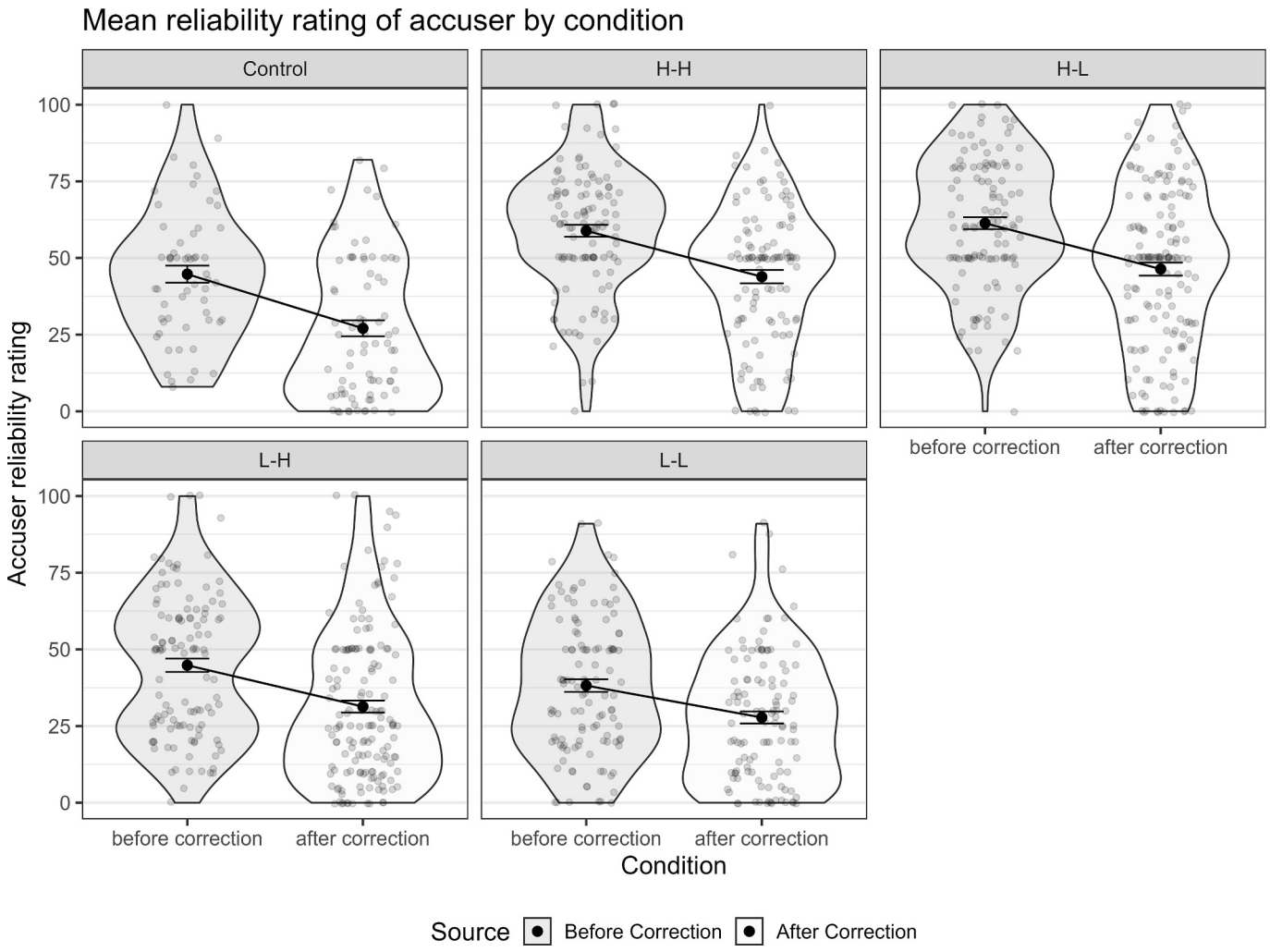


Fig. 7. Mean reliability ratings of the accuser before and after correction, categorised by condition. The magnitude of the decline did not significantly differ between conditions. Note. Error bars represent the standard error of the mean (SE), and jittered points reflect individual participant ratings, illustrating variability within each condition.

likelihood that the bribery occurred. In other words, we expected that stating two aspects of the source's reliability rather than one would emphasise reliability as a relevant component in the evaluation of the claim made. This resulted in four conditions, in addition to reusing the control from Experiment 3.

5.1. Methods

5.1.1. Participants

299 US-based participants were recruited from Prolific, with 60–90[3] participants per condition, mirroring Experiment 3's sample size. Participants' ages ranged from 20 to 83, with a mean of 42.93 (SD = 13.39), evenly distributed between those who identified as female and male. Participants received £0.45 (~$0.57) for their time, with a median completion time of 3.39 min. Six participants were excluded for not meeting the pre-established criterion of correctly answering at least two out of three multiple-choice questions.

5.1.2. Design

The study used a between-subjects design with four conditions: The manipulated variable was the combination of experience and trustworthiness of both prosecutors (see Table 4). The dependent variables remained consistent with Experiment 3 (see https://osf.io/rhjnx).

5.1.3. Materials and procedure

The materials and procedure mirrored Experiment 3, with the only difference being the manipulation of the reliability of the prosecutors, described through both experience and trustworthiness. The aim was to emphasise reliability as a relevant factor in evaluating the likelihood of bribery occurrence.

5.2. Results and discussion

5.2.1. Combined manipulation affects Accuser's reliability ratings

A repeated-measures ANOVA revealed a significant main effect of

[3] In the initial implementation of Experiment 4, unequal sample sizes were introduced as a result of the experiment missing two key conditions in the factorial design. After recognizing the issue, the experiment was rerun with 8 conditions and increased power. Despite the unequal sample sizes between the two runs of the experiment, random assignment was maintained throughout, ensuring unbiased allocation of participants across conditions. To ensure that the unequal sample sizes did not impact the results, an additional analysis was conducted using matched sample sizes for each condition. The results from this analysis were consistent with the original findings and confirmed that the pattern of significant and non-significant results remained unchanged. The supplementary analysis will be made available on OSF for transparency and further examination.

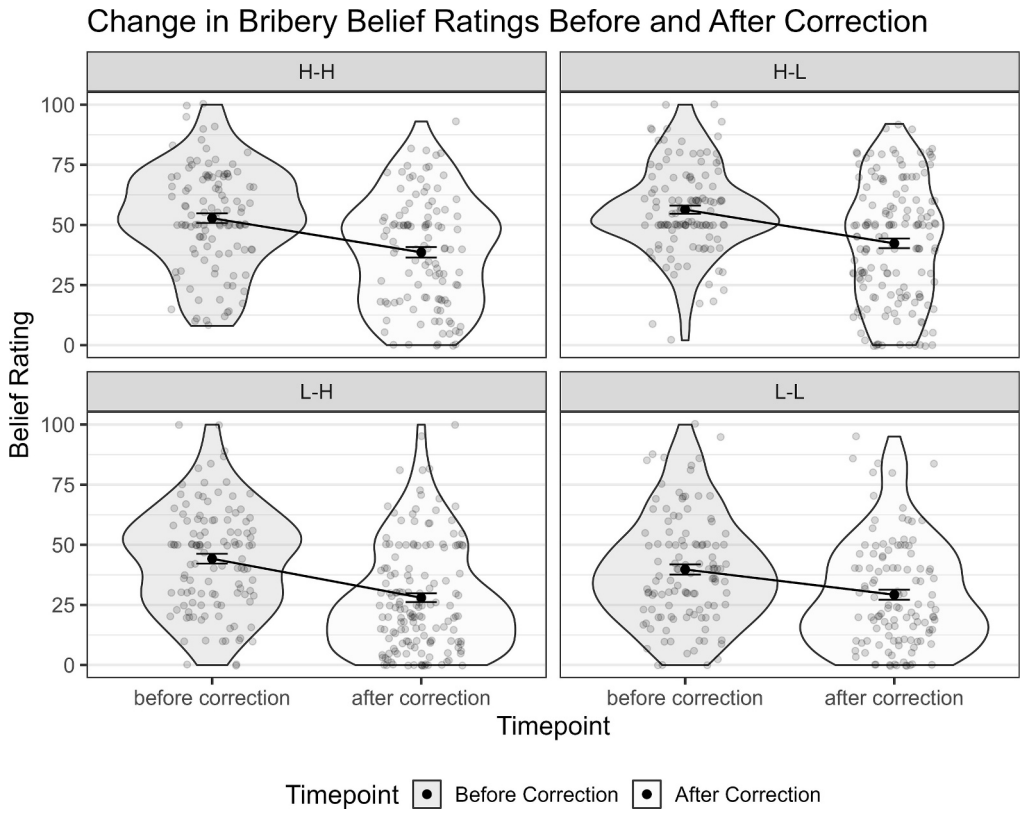


Fig. 8. Mean belief in bribery before and after correction across different conditions. Note. Jittered points represent individual ratings, with larger points indicating mean values. Error bars reflect the standard error of the mean (SE).

Table 4
Conditions for Experiment 4.

Condition	Prosecutor making the allegation	Prosecutor making the retraction
1	Low reliability	Low reliability
2	High reliability	High reliability
3	Low reliability	High reliability
4	High reliability	Low reliability

reliability information about the accuser on participants' reliability ratings ($F(1, 235) = 73.493$, $p < 0.001$). Participants perceived the accuser to be significantly more reliable when they had high experience and high trustworthiness ($M = 58.7$, $SE = 2.15$) compared to low experience and low trustworthiness ($M = 32.6$, $SE = 2.15$) with a mean difference of 25.1 ($SE = 3.04$, $p < 0.001$). (See Fig. 9).

5.2.2. Accuser's reliability affects belief in bribery allegation

A repeated-measures ANOVA showed a significant main effect of the reliability information on the belief that bribery took place ($F(1, 235) = 38.924$, $p < 0.001$). Participants rated the likelihood of bribery significantly higher when the source had high reliability compared to low reliability with a mean difference of 19.2 ($SE = 3.07$, $p < 0.001$). (See Fig. 10).

5.2.3. Participants' perceived reliability of accuser influenced belief in bribery allegation

Linear regression indicated a positive correlation between participants' reliability ratings of the accuser and their belief in bribery occurrence ($\beta = 0.755$, $SE = 0.0345$, $p < 0.001$). The regression model was highly significant ($F(1, 292) = 474.8$, $p < 0.001$), explaining a substantial amount of variance in participants' beliefs ($R^2 = 0.618$). These results underscore the significant influence of individuals' perceptions of the reliability of the source making the bribery allegation on their pre-correction beliefs concerning the bribery.

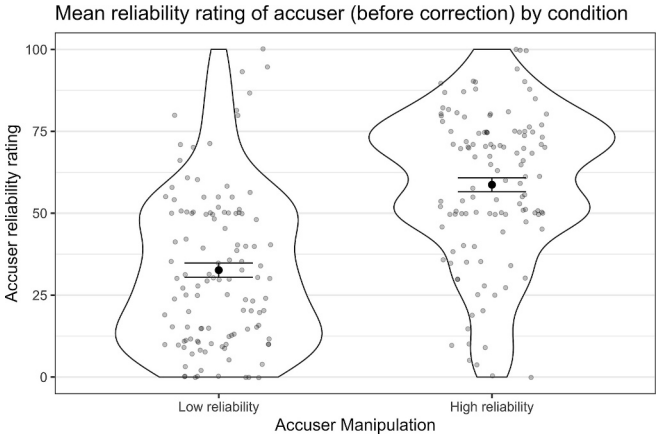


Fig. 9. Mean reliability ratings of the accuser (before correction), categorised by accuser manipulation conditions. Note. Participants rated the accuser manipulated to have higher experience and trustworthiness (High reliability) as more reliable that the accuser manipulated to have lower experience and trustworthiness (Low reliability). Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.

5.3. Judgments at second stage (after correction)

5.3.1. Combined manipulation affects Corrector's reliability ratings

A repeated-measures ANOVA revealed a significant main effect of the corrector's reliability manipulation on participants' perceived reliability of the source ($F(1, 292) = 71.182$, $p < 0.001$). Participants

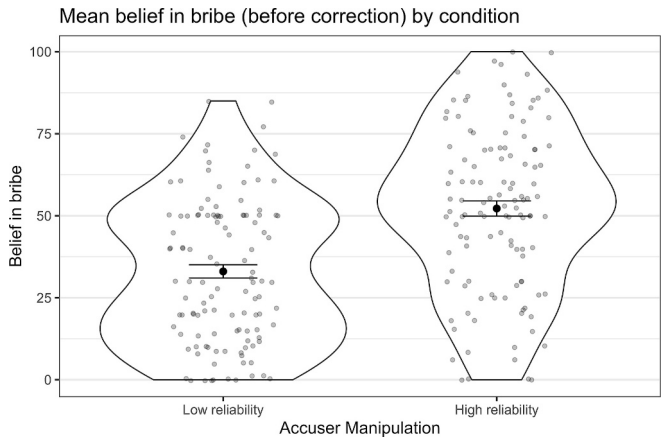


Fig. 10. Mean belief in bribery (before correction), categorised by accuser manipulation conditions. Note. Participants rated the likelihood of bribery significantly higher when the source making the bribery accusation was perceived as having high reliability compared to low reliability. Error bars represent the standard error of the mean (SE). Jittered points illustrate individual ratings, indicating variability within conditions.

perceived the source making the correction as significantly more reliable when both its experience and trustworthiness were high ($M = 69.0$, $SE = 2.06$) compared to low ($M = 44.4$, $SE = 2.06$), with a mean difference of 24.6 ($SE = 2.92$, $p < 0.001$). (See Fig. 11).

5.3.2. *All sources were penalised by contradiction*

Paired sample *t*-tests indicated statistically significant decreases in reliability ratings of the source making the allegation after receiving the correction across all conditions ($p < 0.001$). (See Fig. 12). The magnitude of the drop did not significantly differ across conditions.

5.3.3. *Reliability of both accuser and corrector affects belief in bribery*

A repeated-measures ANOVA was conducted to analyse the effects of time point (before or after correction) within different levels of the first and second sources (either high or low reliability) on participants' beliefs about bribery occurrences (P(Bribe)). (See Fig. 13). The analysis revealed significant main effects for the accuser's ($F(1,232) = 34.14$, $p < 0.001$) and corrector's ($F(1, 232) = 4.01$, $p = 0.046$) reliability information on participants' beliefs regarding the occurrence of bribery. Additionally, a significant main effect of time point was observed ($F(1, 232) = 90.41$, $p < 0.001$), underscoring the impact of the correction on participants' responses. Interaction effects were also examined. While the interaction between first and second source manipulation was not significant ($F(1, 232) = 0.24$, $p = 0.622$), significant interactions emerged between first source and time point ($F(1, 232) = 4.29$, $p = 0.040$), second source and time point ($F(1, 232) = 14.49$, $p < 0.001$), but not for the three-way interaction ($F(1, 232) = 0.39$, $p = 0.531$). These findings suggest that participants' beliefs about bribery are influenced by both the perceived reliability of the individuals making claims and corrections, as well as the presence of a correction itself. (Refer to OSF for further analysis https://osf.io/rhjnx).

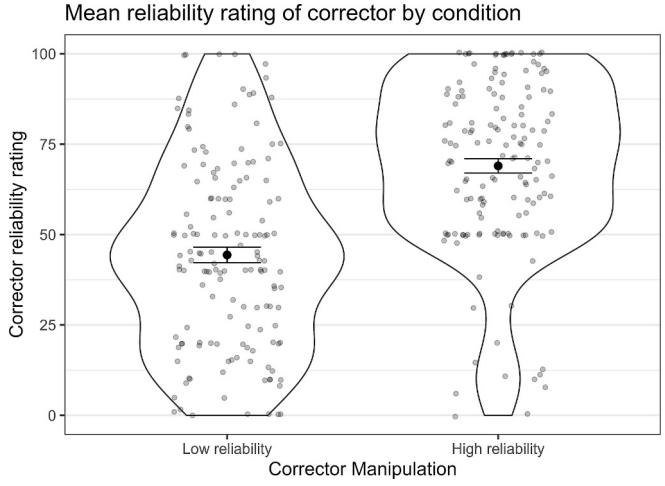


Fig. 11. Mean reliability ratings of the corrector, categorised by accuser manipulation conditions. Note. Participants rated the corrector manipulated to have higher experience and trustworthiness (High reliability) as more reliable that the corrector manipulated to have lower experience and trustworthiness (Low reliability). Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.

6. Qualitative analysis

6.1. *Materials and procedure*

Qualitative data analysis was conducted on responses from an open-text box where participants explained their ratings regarding the likelihood of the bribe. The primary aim was to determine whether participants mentioned the reliability of one or more of the sources making the bribery-related claims (i.e., the accuser and the corrector) in conditions where the source reliability was manipulated, compared to the control condition where reliability was not mentioned. Additionally, the analysis sought to uncover any other justifications participants provided for their ratings.

The underlying assumption for this analysis was that if both sources were perceived as equally reliable, their reliability would be irrelevant for explaining participants' ratings. Consequently, participants in the control condition, where no reliability manipulation occurred, were not expected to mention source reliability in their justifications.

Qualitative data were collected from 205 participants, 117 participants in Experiment 3, 58 participants in Experiment 4, and 30 participants in the control condition. The analysis was conducted following guidance from Dewitt et al. (2024), Varaine (2023), and Aronow et al. (2019).

The hypothesis for the qualitative analysis was that the incidence of participants mentioning source reliability would depend on the strength of the manipulation. Namely, source reliability would be mentioned more frequently in Experiment 4 than in Experiment 3 due to the manipulation involving two pieces of reliability information instead of one. This is because more reliability information is assumed to enhance the salience of the prosecutor's reliability in the context of the probability rating. In line with this assumption, we expect that in the control condition participants would not mention source reliability.

The coding process involved the first coder creating a codebook through exploratory qualitative analysis and then using it to code the data. The process closely followed Dewitt et al.'s protocol for open text boxes (Dewitt et al., 2024). To ensure consistency, a second coder, who was blind to condition and probability ratings, independently provided a second set of codes. For more details on the coding process, refer to the OSF repository (https://osf.io/rhjnx).

6.2. *Results and discussion*

As hypothesised, there were no mentions of reliability in the control condition (see Table 5). This absence aligns with our expectation that without information on source reliability, participants would not consider it in their justifications.

6.2.1. *Comparison between control and experimental conditions*

Chi-squared test results indicated statistically significant differences in the proportion of participants mentioning source reliability between the control group and both experimental conditions. Specifically, the

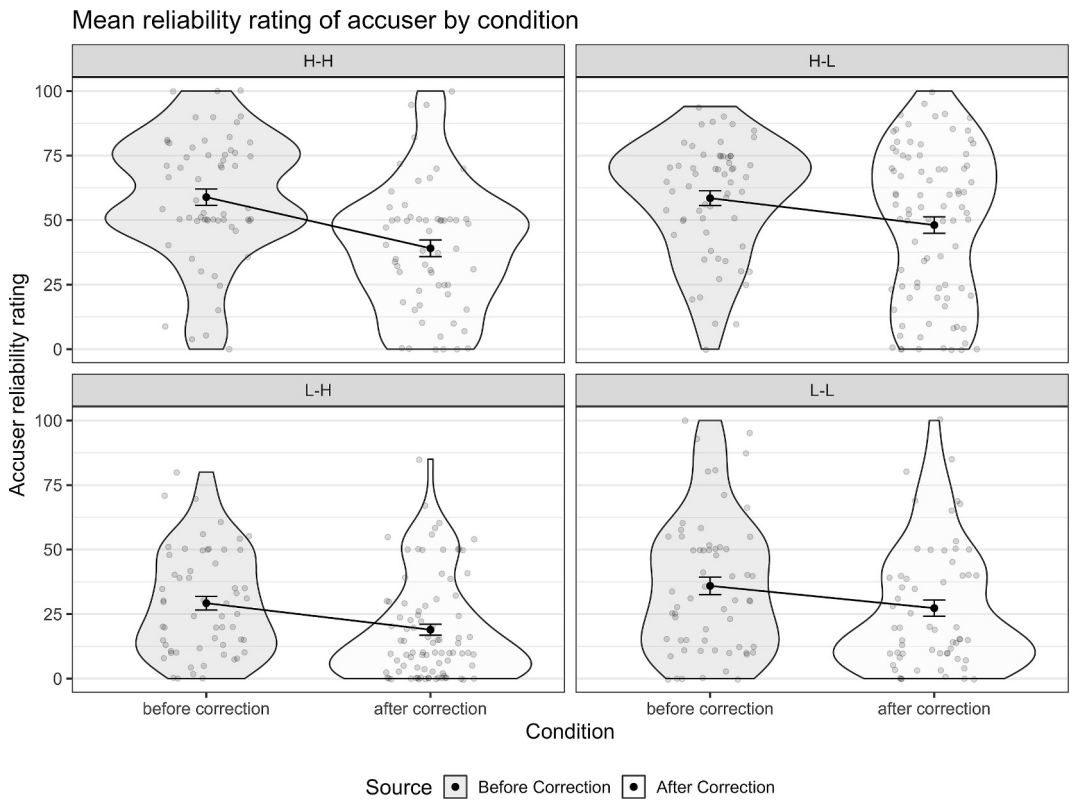


Fig. 12. Mean reliability ratings of the accuser before and after correction, categorised by condition. Note. Paired sample t-tests revealed statistically significant decreases in reliability ratings following the correction across all conditions. The magnitude of the decline did not significantly differ between conditions. Error bars represent the standard error of the mean (SE), and jittered points reflect individual participant ratings, illustrating variability within each condition.

difference between the control group and Experiment 3 was significant ($X^2 = 7.01$, df = 1, $p = 0.008$). Similarly, the difference between the control group and Experiment 4 was significant ($X^2 = 10.67$, df = 1, $p = 0.001$). These results indicate that participants in both experimental conditions, where the reliability of the prosecutor was manipulated, were significantly more likely to mention source reliability compared to participants in the control group. This supports the hypothesis that the introduction of reliability information influences participants to consider and mention source reliability in their justifications.

6.2.2. Comparison between experiment 3 and experiment 4

Within the experimental conditions, mentions of source reliability were more frequent in Experiment 4 (33 %) than in Experiment 3 (23 %). However, the chi-squared test results suggest that this difference is not statistically significant at the 5 % level (X-squared = 1.4095, df = 1, p-value = 0.2351). This indicates that the manipulation involving two pieces of reliability information in Experiment 4 did not result in a significantly higher incidence of participants mentioning source reliability compared to Experiment 3.

While the descriptive data show a trend in the expected direction, the lack of statistical significance might suggest issues with power in the analysis. The smaller effect size detected may require a larger sample size to reach statistical significance. Nonetheless, the data suggest that additional reliability information might increase the salience of source reliability, as indicated by the higher percentage of mentions in Experiment 4.This is further evidenced by the fact that the control condition had no mention of source reliability.

As seen in Table 5 below, a key takeaway of the analysis was that participants mentioned factors other than source reliability when justifying their belief in the briber, such as Henry Light's character and motive. This suggests that individuals rely on a more complex set of heuristics in addition to source reliability when formulating their beliefs.

These results support the hypothesis that participants are more likely to mention source reliability when it is made salient through experimental manipulation. However, while Experiment 4 descriptively shows a higher incidence of reliability mentions compared to Experiment 3, this difference is not statistically significant. Further research with increased sample sizes may be necessary to fully explore the impact of additional reliability information on participants' likelihood of mentioning source reliability. Additionally, it is clear from the analysis that individuals engage in a holistic understanding of the vignette, taking in consideration aspects of the story that go beyond source reliability.

7. General discussion

7.1. Summary of results

In four experiments we investigated how people update their beliefs given contradictory and potentially misleading information about a political candidate taking a bribe. Experiment 1 suggested, in contrast to standard claims in the CIE literature, that people updated their beliefs reasonably. Participants effectively discounted the retracted information, reducing its impact on their beliefs and reverting to their prior baseline beliefs. This does not imply that individuals completely dismissed the possibility of bribery; rather, they returned to their initial level of belief before encountering any evidence of the bribe. We suspect that the lack of a control group in the original Guillory and Geraci (2013) study, in which no allegation is provided, might have led participants' continued belief in the bribe to be attributed to the CIE when they were simply reverting back to baseline beliefs. Experiment 2 further supports our claim, showing that when the source of the retraction was reliable participants discounted the initial allegation, whereas when the

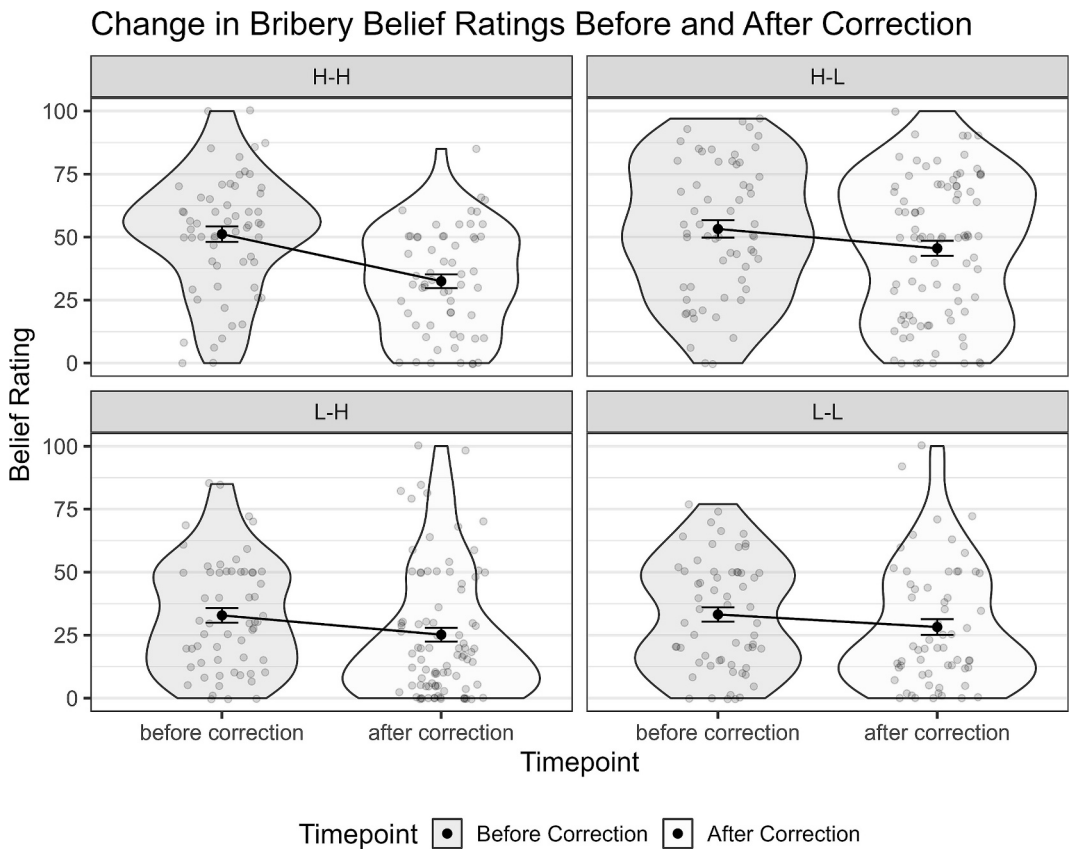


Fig. 13. Mean beliefs about bribery occurrences before and after correction, categorised by condition. Note. Participants' beliefs were significantly influenced by the perceived reliability of both the accuser and the corrector. Jittered points represent individual ratings, while error bars indicate the standard error of the mean (SE).

Table 5
Code description and frequency.

Code	Description	Control	Experiment 3	Experiment 4
Character	The participant states that their perception of Henry Light's character is relevant to justify why they think he did (not) take the bribe	27 %	20 %	21 %
Motive	The participant states that Henry Light's motivation or lack of motivation to commit the bribe is relevant in their probability rating of the bribery occurring	33 %	24 %	33 %
Source reliability	The participant states that the reliability of either the first prosecutor (making the bribery allegation) and/or the second prosecutor (making the correction) are relevant in their perceived likelihood that the bribery took place).	0 %	23 %	33 %
Political prior	The participant states that their prior perception of politicians in general is relevant in their rating of whether the bribery took place	23 %	9 %	5 %

source was unreliable participants did not discount it. These results show that source reliability affects belief updating and that individuals take reliability into account, among other factors, when they evaluate evidence. Experiments 3 and 4 delved deeper into how individuals update their beliefs when confronted with opposing sources of varying reliability. Here our results suggest that individuals are capable of updating their beliefs in a reasonable way and factor in source reliability when it is made salient. Additionally, participants tended to reduce a source's perceived reliability after the source had been contradicted. This happened independently of the reliability of the source making the contradiction. Finally, in the present context we did not find any difference between trustworthiness or expertise in terms of their individual contribution to reliability. This is not to say that a claim of untrustworthiness for an expert source would not reduce their reliability; however, experience and trustworthiness both had a positive effect on perceived reliability. These findings highlight the impact of source expertise when correcting misinformation, which has been discounted in the prior literature (Ecker & Antonio, 2021; Guillory & Geraci, 2013). Overall, our findings reveal the nuances of individual belief updating and contribute to our understanding of the cognitive processes that underpin belief updating.

7.2. *Relation to previous literature*

Our primary findings shed light on the impact of source reliability on belief updating. The present pattern of results is in line with findings by Guillory and Geraci (2013) and Ecker and Antonio (2021), who also show that reliable sources are more effective than unreliable sources at reducing reliance on misinformation, and that these updating processes correspond to behavioural intentions, such as voting. However, our study diverges from Guillory and Geraci (2013) by identifying a reasonable pattern of belief updating in participants. People are capable of completely discounting allegations when these are challenged by highly reliable sources. As a result, the Continued Influence Effect (CIE)

was not observed in this context. Another difference between our findings and those of Guillory and Geraci is that whereas they argue that trustworthiness but not expertise affects whether participants respond to source reliability, in Experiments 3 and 4 we show that people are influenced by both dimensions. This discrepancy might be due to our study using a different definition of expertise. Whereas Guillory and Geraci define expertise as the ability to acquire relevant information, our use of the concept encompassed the more conventional notion of experience. Additionally, our manipulation focused on prosecutors, who, as shown in the norming study, had on average high priors for expertise and trustworthiness. This could constitute a further reason for why expertise in this context was found to have a similar effect as trustworthiness. Based on our studies we do not claim that there is no potential difference between experience and trustworthiness, but we have shown that experience also affects the perceived reliability of a source. It might be that an experienced source with no mention of trustworthiness might by default be perceived as trustworthy, while a trustworthy source would not necessarily be perceived as experienced. Furthermore, to explore whether the present effect would be reduced if the source was experienced but untrustworthy further research should look at how these two dimensions of reliability combine (e.g. Ecker & Antonio, 2021).

Another notable finding in our studies is the effect of contradiction on source reliability. In the final two experiments, the perceived source reliability of the accuser was penalised when their allegation was contradicted regardless of the reliability of the source making the correction. This has widespread implications, especially at a time when the media is awash with contradictory claims from opposing parties, implying that even established, reliable sources can be undermined by less reliable counter-claims. This highlights the importance of further investigating this effect, particularly the extent to which updating beliefs in response to a contradiction can be viewed as a rational process. It could be argued that contradiction in itself can increase one's awareness of alternative possibilities to those elicited by the original source and that therefore downgrading their reliability can be considered reasonable in these circumstances. Introducing a contradictory claim suggests a possibility, however small, that the original source might be incorrect and decreasing their reliability might therefore be rational. In this context, even when an unreliable prosecutor is the source of a contradiction, the chance that they might be correct is not trivial, especially given the fact that a prosecutor is still a reliable figure. Therefore, their claim could provide a valid cue to the perceived unreliability of the accuser. However, this prompts further questions about how to preserve the perceived reliability of an established source, such as the government or traditional news, when faced with such contradictions.

Finally, Experiments 3 and 4 reveal how individuals update their beliefs when faced with varying reliability of two opposing sources. The findings in Experiment 3 fit with Walter and Tukachinsky's (2020) meta-analysis suggesting that the reliability of the source affects people's evaluation of the allegation but not the correction. To further explore this finding, we used a stronger manipulation of reliability in Experiment 4, which showed that source reliability affects both the evaluation of the allegation as well as the correction. Here our findings explain some of the contradictory findings in the literature. Namely, that source reliability becomes a relevant factor in evaluating contradictory claims when it is made sufficiently salient, for example, by presenting information about the sources' experience and trustworthiness. As supported by our qualitative analysis, participants relied on a range of information in addition to source reliability when evaluating the claims, including the character and motive of the accused as well as their political priors. These findings suggest that participants initially relied more on the source's reliability when evaluating the allegation but turned to additional inferred information when faced with two contradictory pieces of evidence. This aligns with the notion, shared by Dias et al. (2020), that people primarily focus on the content of the claim and its alignment with their worldview, with source reliability becoming more relevant afterwards (Lagnado, 2021). When confronted with contradictory information, the complexity of the evaluation increases, prompting participants to consider more context-dependent factors such as the motive and character of the accused. This shifts their focus away from source reliability, which may become less salient as it provides only circumstantial evidence in this context. In sum, when dealing with a more complex evaluation problem, individuals seem to rely primarily on specific situational information to inform their beliefs, rather than focusing solely on the credibility of the source making the claim. When the reliability of the correcting source is made more salient this focus is reduced, suggesting that source reliability might be more contextually relevant. This further supports the notion that source credibility effects may occur only if people actively monitor source credibility (e.g., Sparks & Rapp, 2011; van Boekel et al., 2017).

Another explanation for the findings is that the effect of the reliability of the correcting source might depend on the reliability of the accuser. Zeng et al. (2024) found that the misinformation source moderated the effect of the correction source. When respondents were exposed to misinformation from a highly credible source, their attitudes toward the misinformation were generally positive, regardless of who provided the corrective message. Conversely, when the misinformation came from a less credible source, other social cues, such as the source of the correction, influenced respondents' attitudes. In this context, the correction source acted as a secondary cue, compensating for respondents' lack of confidence in the original misinformation source.

With regards to the CIE in the impression formation literature, our findings address some of the debates in the field. Although our experiments did not focus on impression formation of the politician itself but on belief in a claim, our findings fit with previous literature that supports effective belief updating in impression formation (Ecker & Rodricks, 2020; Mickelberg et al., 2024; Thorson, 2016). For example, Ecker and Rodricks (2020) and Mickelberg et al. (2024) found that when faced with an unequivocal correction, people capably discarded the discredited information. Here the correction was unequivocal as it was factually stated (e.g. "John did not cheat on his wife", Ecker & Rodricks, 2020) as opposed to being a claim from a source within the vignette. A parallel could be drawn here between the unequivocal corrections made in those studies and corrections from highly reliable sources in our study. However, it remains unclear whether the CIE would have been extinguished in those studies if the retraction had been more ambiguous. In line with our present findings we suspect this could be the case. Our paper, therefore supports the notion that individuals update their beliefs effectively after discreditation if they believe in the correction (see also O'Rear & Radvansky, 2020). Further research could explore the effects of the retraction in the present research paradigm on the impression of the political candidate himself - rather than the bribery claim - testing whether, in line with Thorson (2016), individuals might discard the corrected evidence but still reduce their perceived likeability of Henry Light.

7.3. Implications

7.3.1. Is CIE a bias or rational?

Our findings have implications for psychological theorising as well as practical applications. Experiment 4 showed evidence that, when faced with contradictory reports, people seem to intuitively follow the assumptions of the 'rational CIE', appropriately taking into account the reliability of the different sources. In line with Connor Desai et al.'s (2020) re-examination of the CIE, our findings support the notion that individuals are capable of updating their beliefs rationally. These findings also fit with several Bayesian models of belief updating (e.g. Gershman, 2019; Merdes et al., 2021; Zmigrod et al., 2023). Although testing our data against Bayesian network models is needed to provide concrete evidence of rationality, the qualitative pattern of updating shown in Experiment 4 fits with Bayesian prescripts (cf. Shengelia & Lagnado, 2021), and provides a crucial first step toward establishing

that people are updating rationally. This proposal also fits with emerging research advocating for the default assumption of rationality in individuals, reinforcing the idea that people use contextual information and their own reasoning to navigate complex information landscapes effectively (Haselton et al., 2009; Madsen et al., 2024). These implications are crucial for establishing a robust theoretical foundation for psychological research, which can then be translated into effective tools for correcting misinformation.

7.3.2. Broader implications

This paper adds to the current literature by underscoring the potential for rational belief updating in contexts of political impression formation. These findings have notable implications for decision making in politics as well as a range of other applied areas. In the political realm, understanding how to correct misinformation is crucial for maintaining an informed electorate (Pennycook & Rand, 2019). Further studies could also focus on the medical field, where correcting false beliefs, such as misconceptions about the safety of vaccines, is vital to prevent the resurgence of preventable diseases and ensure public health (Loomba et al., 2021). These insights underscore the broader applicability of our research in addressing misinformation across various domains. At this time, a plethora of evidence has been produced claiming that the pervasiveness of misinformation can be read as a tangible threat to the democratic process (e.g., political disengagement and polarisation) (e.g., Friggeri et al., 2014; Lazer et al., 2018; Lewandowsky et al., 2017). Initiatives aimed at alleviating these effects are growing; however, it is essential that these treatments be rooted in exhaustive research and dependable theories.

7.4. Limitations and future directions

Vignette-based research has its limitations. For one, the vignette we used (taken from Guillory & Geraci, 2013) might not generalise to the wider population and real-life political scenarios, which are known to elicit motivated reasoning such as political partisanship (Jern et al., 2014; Jost, 2017). Future research should explore scenarios with higher ecological validity to test the potential limits of the present results. Furthermore, exploring a broader variety of scenarios could provide insights into how the belief updating mechanisms observed in our study applies to different contexts. Specifically, understanding how individuals weigh the reliability of sources and update their beliefs in response to new, conflicting information could be applied to different contexts and fields, such as medical or legal settings, as well as political ones. For instance, how patients update their beliefs about medical treatments or how jurors process conflicting testimonies in court could provide further insights into the robustness and versatility of this belief updating process (Fenton et al., 2013; Mann et al., 2009). Another aspect of the scenario that could be explored in more depth is the motivation of agents to deceive. In the current study, we deliberately used prosecutors to minimize concerns about deceptive intentions, which could potentially influence belief updating. Future research could explore this aspect more directly by using a wider range of sources, allowing for a better understanding of how different agents' motivations to deceive may impact belief revision. It should also be noted that single-item measures, such as the one used in our design, have limitations and could be subject to backfire effects, where corrective information inadvertently strengthens initial beliefs (Swire-Thompson et al., 2022). Although we find this possibility unlikely in the present context, given the nature of our design and the clarity of the corrective information, future studies should aim to replicate these findings with more reliable measures. Another potential limitation of this study is that the sample sizes used in our first two experiments were small (as we followed the sample sizes used in the original study by Guillory and Geraci (2013)). However, given the consistency of our findings with the original results and across all four experiments, we do not believe sample size to be a major concern. Also Experiments 3 and 4 were sufficiently powered, and effectively replicated our key results. In addition, while our manipulation of experience was designed to reflect perceived reliability and was effective in influencing participants' assessments of the source, we acknowledge that experience does not always map directly onto expertise. In some domains, experienced individuals might not possess the most up-to-date knowledge or the highest level of expertise, and future research could explore how different types of experience (e.g., recent vs. outdated) impact perceptions of reliability and expertise.

Finally, more research needs to be conducted on the possibility of individual variability in belief updating. For example, Bayesian network modelling presents a method to model group and individual behaviour, generating comparisons between person-specific Bayesian models and individual inferences. Connor Desai et al. (2020) employed a Bayesian network formalism to explore people's belief updating, showing that they intuitively follow the assumptions needed to classify CIE as a rational process. Their study offers a novel and timely illustration of applying Bayesian networks to the CIE. Future work could apply similar analyses to individual-level updating to capture patterns of individual variability. Moreover, Bayesian networks could also be used to explore how individuals combine different aspects of reliability, such as trustworthiness and expertise, as well as how they integrate claims from both primary and secondary sources (see Fenton et al., 2013, Lagnado, 2021, for application of Bayesian Networks to source reliability questions in the legal domain).

7.5. Conclusion

Misinformation poses a significant challenge to informed decision-making, influencing how individuals form and update their beliefs. The research reported in this paper aims to illuminate the complexities of belief updating mechanisms, enabling more precisely tailored and effective interventions. In this vein our studies suggest that source reliability is a critical factor in how people update their beliefs, and thus highlighting source reliability can be a powerful tool to decrease reliance on misinformation. Our research also suggests that when faced with contradictory evidence people are capable of updating their beliefs rationally. These findings contribute to efforts aimed at mitigating the impact of misinformation and offer insights into the broader literature on belief updating.

Open practices statement

The materials for all experiments are available in the Online Supplementary Material; the data and the supplement are available at https://osf.io/rhjnx. The experiments were preregistered.

Data statement

The data analysed for this paper will be available online.

CRediT authorship contribution statement

Greta Arancia Sanna: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **David Lagnado:** Writing – review & editing, Writing – original draft, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

Special thanks to Henrik Singmann, Victor Btesh, Stephen Dewitt, Amy Rodger and members of the UCL Causal Cognition lab for advice

and feedback. Staff time for the study was part funded by the Economic and Social Research Council's grant to Behavioural Research UK (Grant Ref: ES/Y001044/1).

Appendix A. Experiment 1

Carefully read the following text. You will be asked to remember details of the story and answer some questions.

Vignette.

Message 1: On August 10th/Henry Light announces his campaign for re-election.

Message 2: This is not a surprise to anyone/because he was a good politician/and did many beneficial things for his state during his first term.

Message 3: In his campaign Henry Light promises to improve schools in underprivileged districts. This is very important to him as he was underprivileged growing up.

Message 4: Henry Light embarks on a bus tour around the state to promote his campaign. During this time he meets with many of the local citizens.

Message 5: There are many secret meetings scheduled during the campaign. No media are allowed in during the meetings.

Message 6a (No Allegation): In the middle of his campaign, it is reported that Henry Light will schedule a public debate with his opponent.

Message 6b (Allegation without correction): In the middle of his campaign, it is reported that Henry Light was seen taking bribe money.

Message 6c (Allegation with correction): In the middle of his campaign, it is reported that Henry Light was seen taking bribe money.

Message 7: Henry Light was raised by a single mother, with his 4 brothers and 3 sisters.

Message 8: From a young age Henry Light vowed that he would be successful as an adult, and never be poor again. He wanted to give his mother and siblings what they never had growing up.

Message 9: Henry Light takes time away from his campaign to take a vacation with his wife and children. During this time he does not speak with the media.

Message 10: Henry Light and his opponent meet to have a public debate about their stand on important issues. The debate is televised on the local media station.

Message 11: Two months into his campaign/Henry makes a sizeable donation to the homeless.

Message 12a (No Allegation): The local school children follow the election coverage.

Message 12b (Allegation without correction): The local school children follow the election coverage.

Message 12c (Allegation with correction): A source reports a correction that the previous report was incorrect and Henry Light did not take any bribe money.

Message 13: Election day arrives on May 16th.

Factual questions (multiple choice)

1. Why does Henry Light promise to improve schools in underprivileged districts?
   a. He wants his children to have a good education.
   b. He was underprivileged growing up.
   c. He wants to support the homeless community.
   d. He didn't go to school but wished he had.
2. What did Henry Light vow when he was younger?
   a. That he would be successful as an adult, and never be poor again.
   b. That he would one day become president.
   c. That he would eradicate homelessness.
   d. That he would spend all his free time with his family.
3. Who did Henry Light take a vacation with during his campaign?
   a. His campaign manager.
   b. Members of his staff.
   c. His wife and children.
   d. He went on his own.

Base rating/CIE.

How likely is Henry Light to have taken a bribe during his election campaign? (scale of 1–100).

Control 2.

How reliable is the source that reported that Henry Light was seen taking bribe money? (scale of 1–100).

Appendix B. Experiment 2: Vignette variations from Experiment 1

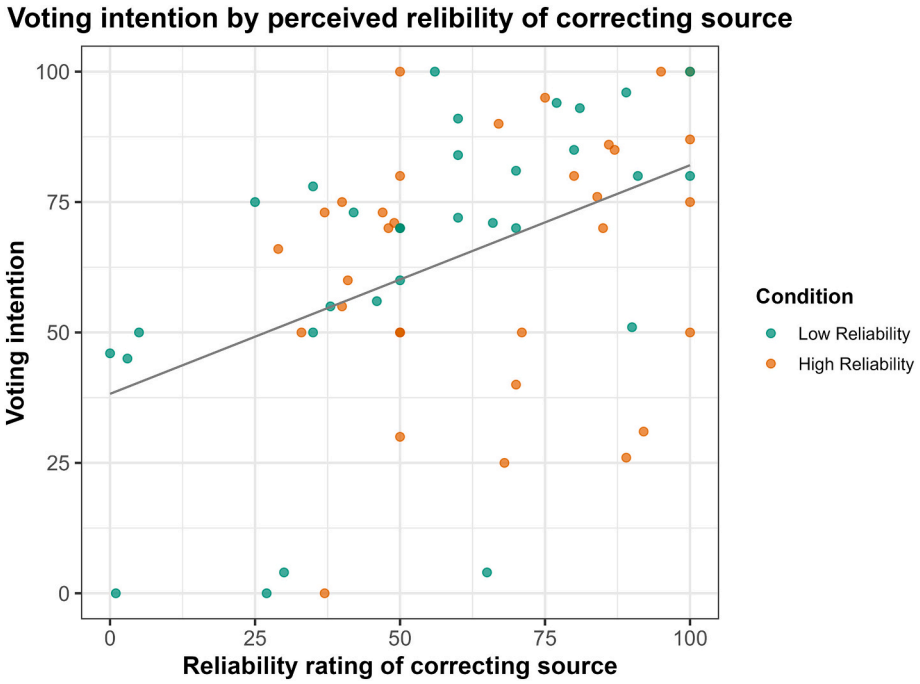
Correction conditions.

Message 12a: The local school children follow the election coverage.

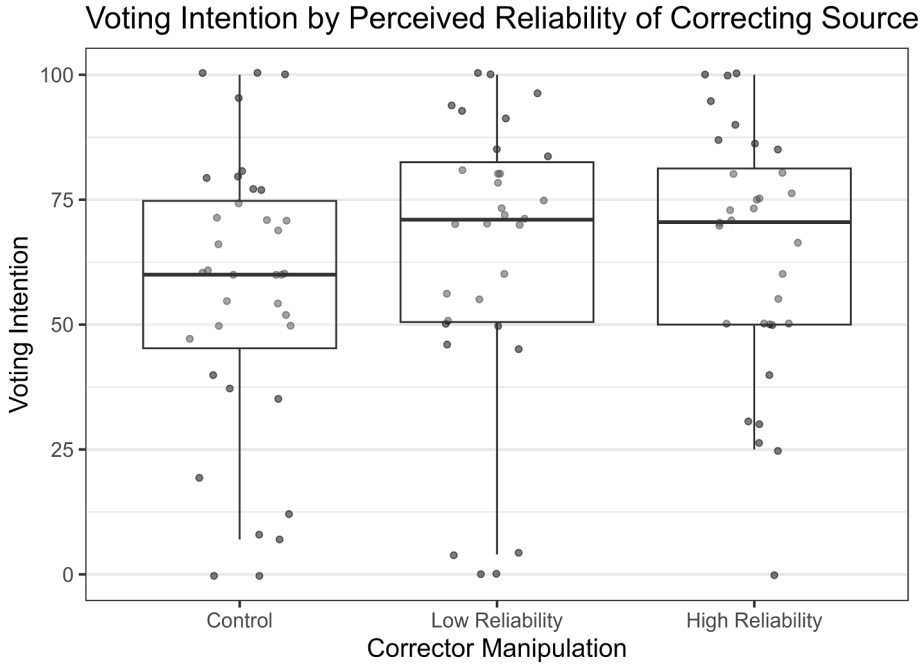
Message 12b (Low expertise, low trustworthiness): A celebrity/political satire news channel/The wife of Henry Light's opponent reports a correction that the previous report was incorrect/and Henry Light did not take any bribe money.

Message 12c (High expertise, high trustworthiness): A government report/a representative from the state legal department/A district attorney reports a correction that the previous report was incorrect/and Henry Light did not take any bribe money.

Appendix C. Appendix



Voting intention by reliability rating split by correcting source. Note. The horizontal axis indicates the reliability rating of the corrector, while the vertical axis displays the voting intention. Ratings for reliability ranged from 0 (very unreliable) to 100 (very reliable) while ratings for voting intention ranged from 0(very unlikely) to 100 (very likely). Participants who rate the reliability of the correcting source more highly also exhibited higher intention to vote for the candidate.



Voting intention by corrector manipulation. Note. The horizontal axis indicates the manipulation of the corrector, while the vertical axis displays the voting intention.

Further ANOVA analysis explored how the reliability manipulation of the correction affected voting intention. The mean voting intentions for the Control condition, Low Reliability, and High Reliability conditions were 56.6 (SE = 4.53), 64.0 (SE = 4.89), and 64.7 (SE = 4.81), respectively. Pairwise comparisons showed no statistically significant differences between any of the conditions: Control vs. Low Reliability (estimate = −7.39, SE

= 6.67, $p = 0.5113$, p = 0.5113, p = 0.5113), Control vs. High Reliability (estimate = −8.05, SE = 6.61, $p = 0.446$, p = 0.446, 1p = 0.446), and Low Reliability vs. High Reliability (estimate = −0.66, SE = 6.86, $p = 0.995$, p = 0.995, p = 0.995). These results indicate that voting intention did not significantly vary with the corrector manipulation. This suggests that voting intention may be influenced more by other factors, such as individual perceptions of the corrector's reliability, rather than the experimental manipulation itself.

Appendix D. Experiment 3 - Vignette variations from Experiment 1 and 2

All conditions.
Control

- **Message 6**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money.
- **Message 12:** a different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money.

Low vs Low experience

- **Message 6a**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has no experience investigating corruption.
- **Message 12a:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor also has no experience investigating corruption.

High vs High experience

- **Message 6b**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has extensive experience investigating corruption.
- **Message 12b:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor also has extensive experience investigating corruption.

Low vs high experience

- **Message 6c**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has no experience investigating corruption.
- **Message 12c:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has extensive experience investigating corruption.

High vs low experience

- **Message 6d**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has extensive experience investigating corruption.
- **Message 12d:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has no experience investigating corruption.

Not Trustworthy vs not trustworthy

- **Message 6e**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has allegations of previous wrongdoings against him.
- **Message 12e:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor also has allegations of previous wrongdoings against him.

Trustworthy vs trustworthy

- **Message 6f**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has a clean record.
- **Message 12f:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor also has a clean record.

Trustworthy vs not trustworthy

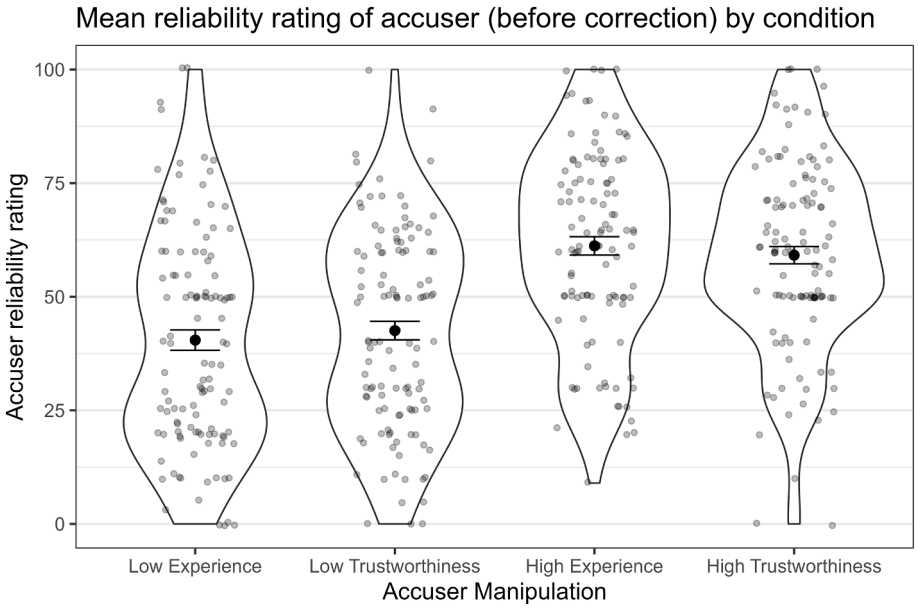
- **Message 6 g**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has a clean record.
- **Message 12 g:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has allegations of previous wrongdoings against him.

Not Trustworthy vs trustworthy

- **Message 6 h**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has allegations of previous wrongdoings against him.

- **Message 12 h:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has a clean record.

Appendix E



Mean reliability ratings of the accuser (before correction), categorised by accuser manipulation conditions including both experience and trustworthiness manipulations. Note. Participants rated the accuser manipulated to have higher experience and trustworthiness as more reliable that the accuser manipulated to have lower experience and trustworthiness. Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.

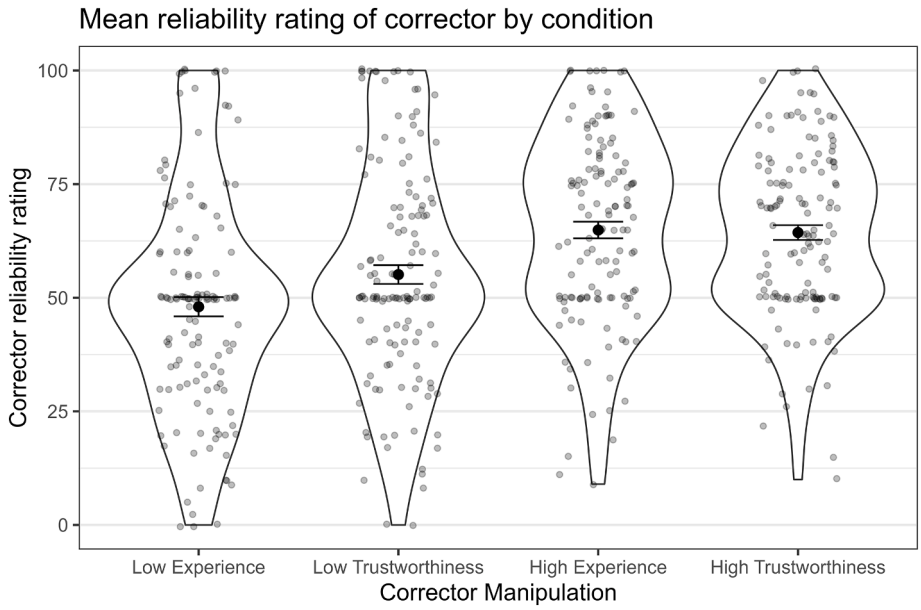
An ANOVA was conducted to compare the mean differences between various conditions of the accuser reliability manipulation to determine if there were statistically significant differences in perceived reliability, specifically focusing on "Low Experience" vs. "Low Trustworthiness" and "High Experience" vs. "High Trustworthiness."

Pairwise Comparisons of Interest

- **Low Experience vs. Low Trustworthiness**: The estimated difference between these conditions was −1.99 (SE = 2.80), with a t-ratio of −0.714 and a *p*-value of 0.953. This non-significant result suggests no meaningful difference in perceived reliability between the *Low Experience* and *Low Trustworthiness* conditions.
- **High Experience vs. High Trustworthiness**: The estimated difference was −4.78 (SE = 2.83), with a t-ratio of −1.69 and a p-value of 0.441. This result also indicates no statistically significant difference in perceived reliability between the *High Experience* and *High Trustworthiness* conditions.

E.1. Conclusion

The pairwise comparisons show that there is no statistically significant difference in perceived reliability between the *Low Experience* and *Low Trustworthiness* conditions, nor between the *High Experience* and *High Trustworthiness* conditions. These findings suggest that participants perceived accusers with low and high levels of expertise and trustworthiness similarly when evaluating their reliability.



Mean reliability ratings of the corrector, categorised by accuser manipulation conditions. *Participants rated the corrector manipulated to have higher experience and trustworthiness (High reliability) as more reliable that the corrector manipulated to have lower experience and trustworthiness (Low reliability). Error bars represent the standard error of the mean (SE). Jittered points indicate individual ratings, reflecting variability within conditions.*

An ANOVA was conducted to compare the mean differences between various conditions of the corrector reliability manipulation to determine if there were statistically significant differences in perceived reliability, specifically focusing on "Low Experience" vs. "Low Trustworthiness" and "High Experience" vs. "High Trustworthiness."

Pairwise Comparisons Results.

The pairwise comparisons of the corrector reliability manipulation variable, adjusted using the Tukey method for multiple comparisons, revealed the following results:

- **Low Experience vs. Low Trustworthiness**: The estimated difference was −7.09 (SE = 2.73, df = 606), with a t-ratio of −2.59 and a p-value of 0.073. This suggests no statistically significant difference between these conditions at the conventional alpha level ($p > 0.05$).
- **High Experience vs. High Trustworthiness**: The estimated difference was 0.557 (SE = 2.76, df = 606), with a t-ratio of 0.202 and a p-value of 0.999. This indicates no statistically significant difference between "High Experience" and "High Trustworthiness."

E.2. Conclusion

The pairwise comparison analysis demonstrates that there is no statistically significant difference between the "Low Experience" and "Low Trustworthiness" conditions or between the "High Experience" and "High Trustworthiness" conditions. These findings suggest that participants did not perceive a meaningful distinction in reliability between these specific levels of experience and trustworthiness in the context of the corrector reliability manipulation variable. However, given the proximity of the p-value for the difference between "Low Experience" and "Low Trustworthiness" ($p = 0.073$) to the significance threshold, future research could explore this relationship more closely to determine if there is a meaningful distinction under different conditions or with larger sample sizes.

Appendix F. Experiment 4 - Vignette variations from Experiment 1, 2 and 3

Not trustworthy AND low competence vs Not trustworthy AND low competence

- **Message 6i**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has no experience investigating corruption, and he also has allegations of previous wrongdoings against him.
- **Message 12i:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has no experience investigating corruption, and he also has allegations of previous wrongdoings against him.

Trustworthy AND high competence vs Trustworthy AND high competence

- **Message 6j**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has extensive experience investigating corruption and a clean record.
- **Message 12j:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has extensive experience investigating corruption and a clean record.

Not trustworthy AND low competence vs Trustworthy AND high competence

- **Message 6 k**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has no experience investigating corruption, and he also has allegations of previous wrongdoings against him.
- **Message 12 k:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has extensive experience investigating corruption and a clean record.

Not trustworthy AND high competence vs Trustworthy and low competence

- **Message 6 l**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has extensive experience investigating corruption, but he also has allegations of previous wrongdoings against him.
- **Message 12 l:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has no experience investigating corruption but a clean record.

Trustworthy AND high competence vs Not trustworthy AND low competence

- **Message 6 m**: In the middle of his campaign, a prosecutor stated that Henry Light was seen taking bribe money. This prosecutor has extensive experience investigating corruption, and he also has a clean record.
- **Message 12 m:** A different prosecutor reports that the previous report was incorrect and Henry Light did not take any bribe money. This prosecutor has no experience investigating corruption and also has allegations of previous wrongdoings against him.

Appendix G. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2025.106090.

Data availability

Anonymised research data and supplementary material are available on OSF through the following link: https://osf.io/rhjnx

References

Ames, D. R., & Kammrath, L. K. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*(3), 187–209.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*(6), 1037–1049. https://doi.org/10.1037/h0077720

Aronow, P. M., Baron, J., & Pinson, L. (2019). A note on dropping experimental subjects who fail a manipulation check. *Political Analysis, 27*, 572–589.

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5*(1), 1–21.

Bandura, A. (1977). *Social Learning Theory, Englewood Cliffs, NJ*. Prentice Hall.

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review, 37*(1), 38–54.

van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition, 45*, 168–181. https://doi.org/10.3758/s13421-016-0649-0

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology, Oxford*. Oxford University Press.

Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology, 20*(1), 49–96. https://doi.org/10.1080/10463280802643640

Brydges, C. R., Gignac, G. E., & Ecker, U. K. H. (2018). Working memory capacity, short-term memory capacity, and the continued influence effect: A latent-variable analysis. *Intelligence, 69*, 117–122. https://doi.org/10.1016/j.intell.2018.03.009

Chan, M. P. S., & Albarracín, D. (2023). A meta-analysis of correction effects in science-relevant misinformation. *Nature Human Behaviour, 7*(9), 1514–1525.

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A Meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*(11), 1531–1546. https://doi.org/10.1177/0956797617714579

Cobb, M. D., Nyhan, B., & Reifler, J. (2013). Beliefs Don't always persevere: How political figures are punished when positive information about them is discredited. *Political Psychology, 34*(3), 307–326. https://doi.org/10.1111/j.1467-9221.2012.00935.x

Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences, 116*(20), 9802–9807. https://doi.org/10.1073/pnas.1903222116

Connor Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for web-based measures of the continued-influence effect. *Behavior Research Methods, 51*(3), 1426–1440. https://doi.org/10.3758/s13428-018-1066-z

Connor Desai, S. A., Pilditch, T. D., & Madsen, J. K. (2020). The rational continued influence of misinformation. *Cognition, 205*, Article 104453. https://doi.org/10.1016/j.cognition.2020.104453

De keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence, 65*, 107–110. https://doi.org/10.1016/j.intell.2017.10.005

Dewitt, S., Liefgreen, A., Adler, N., & Strittmatter, L. E. (2024). *Analysing Open Text Box Data: A Pragmatic and Reflexive Guide*. https://doi.org/10.31234/osf.io/7qsng

Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review.*. https://doi.org/10.37016/mr-2020-001

Dourado, G. B., Volpato, G. H., de Almeida-Pedrin, R. R., Oltramari, P. V. P., Fernandes, T. M. F., & Conti, A. C. D. C. F. (2021). Likert scale vs visual analog scale for assessing facial pleasantness. *American Journal of Orthodontics and Dentofacial Orthopedics, 160*(6), 844–852.

Ecker, U. K., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition, 49*, 631–644.

Ecker, U. K., 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.

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition, 6*(2), 185–192. https://doi.org/10.1037/h0101809

Ecker, U. K. H., Lewandowsky, S., & Apai, J. (2011). Terrorists brought down the plane!—No, actually it was a technical fault: Processing corrections of emotive information. *Quarterly Journal of Experimental Psychology, 64*(2), 283–310. https://doi.org/10.1080/17470218.2010.497927

Ecker, U. K. H., Lewandowsky, S., Cheung, C. S. C., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language, 85*(3758), 101–115. https://doi.org/10.1016/j.jml.2015.09.002

Ecker, U. K. H., Prike, T., Paver, A. B., Scott, R. J., & Swire-Thompson, B. (2024). Don't believe them! Reducing misinformation influence through source discreditation. *Cognitive Research: Principles and Implications, 9*(1), 52. https://doi.org/10.1186/s41235-024-00581-7

Ecker, U. K. H., & Rodricks, A. E. (2020). Do false allegations persist? Retracted misinformation does not continue to influence explicit person impressions. *Journal of Applied Research in Memory and Cognition, 9*(4), 587–601. https://doi.org/10.1016/j.jarmac.2020.08.003

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. *Cognitive Science, 37*(1), 61–102. https://doi.org/10.1111/cogs.12004

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Vol. 23. Advances in Experimental Social Psychology* (pp. 1–74). Academic Press.

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In , *Vol. 8. Proceedings of the International AAAI Conference on Web and Social Media* (pp. 101–110).

Garett, R., & Young, S. D. (2021). Online misinformation and vaccine hesitancy. *Translational Behavioral Medicine, 11*(12), 2194–2199.
Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review, 26*(1), 13–28. https://doi.org/10.3758/s13423-018-1488-8
Gordon, A., Brooks, J. C. W., Quadflieg, S., Ecker, U. K. H., & Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing. *Neuropsychologia, 106*, 216–224. https://doi.org/10.1016/j.neuropsychologia.2017.10.003
Gordon, A., Ecker, U. K. H., & Lewandowsky, S. (2019). Polarity and attitude effects in the continued-influence paradigm. *Journal of Memory and Language, 108*. https://doi.org/10.1016/j.jml.2019.104028
Gradoń, K. (2020). Crime in the time of the plague: Fake news pandemic and the challenges to law-enforcement and INTELLIGENCE community. *Society Register, 4*(2), 133–148. https://doi.org/10.14746/sr.2020.4.2.10
Guardian. (2024). *Russia Propaganda Group Behind Fake Kamala Harris Hit-and-Run Story, Says Microsoft*. Guardian. Available at https://theguardian.com/us-news/2024/sep/18/kamala-harris-fake-hit-run-story-russia-propaganda-storm-1516 [Accessed 4 Nov. 2024].
Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. *Psychonomic Bulletin & Review, 17*(1), 73–81. https://doi.org/10.3758/PBR.17.1.73
Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition, 2*(4), 201–209. https://doi.org/10.1016/j.jarmac.2013.10.001
Harris, A. J. L., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1366–1373. https://doi.org/10.1037/a0016567
Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science, 40*(6), 1496–1533. https://doi.org/10.1111/cogs.12276
Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition, 27*(5), 733–763. https://doi.org/10.1521/soco.2009.27.5.733
Horai, J., Naccari, N., & Fatoullah, E. (1974). The effects of expertise and physical attractiveness upon opinion agreement and liking. *Sociometry, 3*(7), 601–606.
Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness*. *Public Opinion Quarterly, 15*(4), 635–650. https://doi.org/10.1086/266350
Jardina, A., & Traugott, M. (2019). The genesis of the birther rumor: Partisanship, racial attitudes, and political knowledge. *Journal of Race, Ethnicity, and Politics, 4*(1), 60–80. https://doi.org/10.1017/rep.201 8.25
Jarvstad, A., & Hahn, U. (2011). Source reliability and the conjunction fallacy. *Cognitive Science, 35*(4), 682–711. https://doi.org/10.1111/j.1551-6709.2011.01170.x
Jern, A., Chang, K. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review, 121*(2), 206–224. https://doi.org/10.1037/a0035941
Johnson, H. H., & Izzett, R. R. (1969). Relationship between authoritarianism and attitude change as a function of source credibility and type of communication. *Journal of Personality and Social Psychology, 13*(4), 317–321. https://doi.org/10.1037/h0028440
Johnson, H. H., Torvicia, J., & Poprick, M. (1968). Effects of source credibility on the relationship between authoritarianism and attitude change. *Journal of Personality and Social Psychology, 9*(179–1), 83.
Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420
Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. *Political Psychology, 38*(2), 167–208. https://doi.org/10.1111/pops.12407
Kelman, H. C., & Hovland, C. I. (1953). "Reinstatement" of the communicator in delayed measurement of opinion change. *The Journal of Abnormal and Social Psychology, 48*(3), 327–335. https://doi.org/10.1037/h0061861
Kendeou, P., Butterfuss, R., Kim, J., & Van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition, 47*(1), 33–46. https://doi.org/10.3758/s13421-018-0848-y
Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes, 51*(5–6), 374–397. https://doi.org/10.1080/0163853X.2014.913961
Kumkale, G. T., Albarracín, D., & Seignourel, P. J. (2010). The effects of source credibility in the presence or absence of prior attitudes: Implications for the Design of Persuasive Communication Campaigns. *Journal of Applied Social Psychology, 40*(6), 1325–1356. https://doi.org/10.1111/j.1559-1816.2010.00620.x
Lagnado, D. (2021). *Explaining the Evidence: How the Mind Investigates the World*. Cambridge University Press.
Lazer, D. M., 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.
Lee, S. K., Sun, J., Jang, S., & Connelly, S. (2022). Misinformation of COVID-19 vaccines and vaccine hesitancy. *Scientific Reports, 12*(1), 13681.
Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful Debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131. https://doi.org/10.1177/1529100612451018
Lirtzman, S. I., & Shuv-Ami, A. (1986). Credibility of source of communication on products' safety hazards. *Psychological Reports, 58*(707–7), 18.
Loomba, S., De Figueiredo, A., Piatek, S. J., De Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour, 5*(3), 337–348.
MacFarlane, D., Tay, L. Q., Hurlstone, M. J., & Ecker, U. K. H. (2021). Refuting spurious COVID-19 treatment claims reduces demand and misinformation sharing. *Journal of Applied Research in Memory and Cognition, 10*(2), 248–258. https://doi.org/10.1037/h0101793
Maddux, J., & Rogers, R. (1980). Effects of source expertness, physical attractiveness, and supporting arguments on persuasion: A case of brains over beauty. *Journal of Personality and Social Psychology, 39*, 235–244. https://doi.org/10.1037/0022-3514.39.2.235
Madsen, J. K. (2016). Trump supported it?! A Bayesian source credibility model applied to appeals tospecific American presidential candidates' opinions. *Proceedings of the Annual Meeting of the Cognitive Science Society, 38*.
Madsen, J. K., de-Wit, L., Ayton, P., Brick, C., de-Moliere, L., & Groom, C. J. (2024). Behavioral science should start by assuming people are reasonable. *Trends in Cognitive Sciences, 28*(7), 583–585. https://doi.org/10.1016/j.tics.2024.04.010
Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source dependence on belief and reliability revision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(9), 1795–1805. https://doi.org/10.1037/xlm0000846
Mann, D. M., Ponieman, D., Leventhal, H., & Halm, E. A. (2009). Predictors of adherence to diabetes medications: The role of disease and medication beliefs. *Journal of Behavioral Medicine, 32*, 278–284.
Merdes, C., von Sydow, M., & Hahn, U. (2021). Formal models of source reliability. *Synthese, 198*(23), 5773–5801. https://doi.org/10.1007/s11229-020-02595-2
Mickelberg, A., Walker, B., Ecker, U., Howe, P. D. L., Perfors, A., & Fay, N. (2024). Does mud really stick? No evidence for continued influence of misinformation on newly formed person impressions. *Collabra: Psychology, 10*(1), Article 92332. https://doi.org/10.1525/collabra.92332
Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*, 303–330.
Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarization. In F. Zenker (Ed.), *Bayesian Argumentation, The Practical Side of Probability* (pp. 113–133). Dordrecht: Springer.
O'Rear, A. E., & Radvansky, G. A. (2020). Failure to accept retractions: A contribution to the continued influence effect. *Memory & Cognition, 48*(1), 127–144. https://doi.org/10.3758/s13421-019-00967-9
Oyserman, D., & Dawson, A. (2020). Your fake news, our facts: Identity-based motivation shapes what we believe, share, and accept. In *The Psychology of Fake News* (pp. 173–195). Routledge.
Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., … Ecker, U. (2019). Evaluation of a template for countering misinformation—Real-world autism treatment myth debunking. *PLoS ONE, 14*(1), Article e0210746. https://doi.org/10.1371/journal.pone.0210746
Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences, 116*(7), 2521–2526.
Petratos, P. N. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons, 64*(6), 763–774. https://doi.org/10.1016/j.bushor.2021.07.012
Pilgrim, C., Sanborn, A., Malthouse, E., & Hills, T. T. (2024). Confirmation bias emerges from an approximation to Bayesian reasoning. *Cognition, 245*, Article 105693. https://doi.org/10.1016/j.cognition.2023.105693
Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five Decades' evidence. *Journal of Applied Social Psychology, 34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x
Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition, 42*(1), 11–26. https://doi.org/10.3758/s13421-013-0339-0
Ribeiro, M. H., Calais, P. H., Almeida, V. A. F., & Meira, W., Jr. (2017). "Everything I Disagree With Is #FakeNews": Correlating Political Polarization and Spread of Misinformation (arXiv:1706.05924). arXiv http://arxiv.org/abs/1706.05924.
Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 62–74. https://doi.org/10.1037/xlm0000155
Ross, A. S., & Rivers, D. J. (2018). Discursive deflection: Accusation of "fake news" and the spread of mis-and disinformation in the tweets of president trump. *Social Media+ Society, 4*(2), Article 2056305118776010.
Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*(5), 880–892. https://doi.org/10.1037/0022-3514.32.5.880
Shengelia, T., & Lagnado, D. (2021). Are jurors intuitive statisticians? Bayesian causal reasoning in legal contexts. *Frontiers in Psychology, 11*, Article 519262.
Smith, V. L., & Ellsworth, P. C. (1987). The social psychology of eyewitness accuracy: Misleading questions and communicator expertise. *Journal of Applied Psychology, 72*(2), 294–300. https://doi.org/10.1037/0021-9010.72.2.294
Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 37*, 230–247. https://doi.org/10.1037/a0021331
Steblay, N., Hosch, H. M., Culhane, S. E., & McWethy, A. (2006). The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: A meta-analysis.

Law and Human Behavior, 30(4), 469–492. https://doi.org/10.1007/s10979-006-9039-7

Suarez-Lledo, V., & Alvarez-Galvez, J. (2021). Prevalence of health misinformation on social media: Systematic review. *Journal of Medical Internet Research, 23*(1), Article e17187.

Sung, Y. T., & Wu, J. S. (2018). The visual analogue scale for rating, ranking and paired-comparison (VAS-RRP): A new technique for psychological measurement. *Behavior Research Methods, 50*, 1694–1715.

Susmann, M. W., & Wegener, D. T. (2022). The role of discomfort in the continued influence effect of misinformation. *Memory & Cognition, 50*(2), 435–448.

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1948–1961. https://doi.org/10.1037/xlm0000422

Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: Challenges and recommendations. *Annual Review of Public Health, 41*(1), 433–451.

Swire-Thompson, B., Miklaucic, N., Wihbey, J. P., Lazer, D., & DeGutis, J. (2022). The backfire effect after correcting misinformation is strongly associated with reliability. *Journal of Experimental Psychology. General, 151*(7), 1655–1665. https://doi.org/10.1037/xge0001131

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*. https://doi.org/10.1080/10584609.2015.1102187

Uleman, J. S., & Kressel, L. M. (2013). A Brief History of Theory and Research on Impression Formation. In *Oxford Handbook of Social Cognition* (pp. 53–73).

Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges, 1*(2), Article 1600008.

Varaine, S. (2023). How dropping subjects who failed manipulation checks can bias your results: An illustrative case. *Journal of Experimental Political Science, 10*(2), 299–305.

Waldman, A. E. (2017). The marketplace of fake news symposium: Hate Crime v. hate speech: Exploring the first amendment. *University of Pennsylvania Journal of Constitutional Law, 20*(4), 845–870.

Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs, 85*(3), 423–441. https://doi.org/10.1080/03637751.2018.1467564

Walter, N., & Tukachinsky, R. (2020). A Meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research, 47*(2), 155–177. https://doi.org/10.1177/0093650219854600

Weeks, B. E., & Garrett, R. K. (2014). Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 U.S. presidential election. *International Journal of Public Opinion Research, 26*(4), 401–422. https://doi.org/10.1093/ijpor/edu005

Whittaker, J. O., & Meade, R. D. (1968). Retention of opinion change as a function of differential source credibility. *International Journal of Psychology, 3*, 103–108.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *The Quarterly Journal of Experimental Psychology Section A, 40*(2), 361–387. https://doi.org/10.1080/02724988843000168

Wilkes, A. L., & Reynolds, D. J. (1999). On certain limitations accompanying Readers' interpretations of corrections in episodic text. *The Quarterly Journal of Experimental Psychology Section A, 52*(1), 165–183. https://doi.org/10.1080/713755808

World Economic Forum. (2024). *Global Risks Report 2024*. World Economic Forum. https://www.weforum.org/publications/global-risks-report-2024/.

Zeng, H.-K., Lo, S.-Y., & Li, S.-C. S. (2024). Credibility of misinformation source moderates the effectiveness of corrective messages on social media. *Public Understanding of Science, 33*(5), 587–603. https://doi.org/10.1177/09636625231215979

Zhu, B., Chen, C., Loftus, F., & E., Lin, C., & Dong, Q.. (2010). Treat and trick: A new way to increase false memory. *Applied Cognitive Psychology, 24*(9), 1199–1208. https://doi.org/10.1002/acp.1637

Zmigrod, L., Burnell, R., & Hameleers, M. (2023). The misinformation receptivity framework: Political misinformation and disinformation as cognitive Bayesian inference problems. *European Psychologist, 28*(3), 173–188. https://doi.org/10.1027/1016-9040/a000498