



Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect

Ullrich K. H. Ecker¹ · Luke M. Antonio¹

Accepted: 10 December 2020 / Published online: 15 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

The continued influence effect refers to the finding that people often continue to rely on misinformation in their reasoning even if the information has been retracted. The present study aimed to investigate the extent to which the effectiveness of a retraction is determined by its credibility. In particular, we aimed to scrutinize previous findings suggesting that perceived trustworthiness but not perceived expertise of the retraction source determines a retraction's effectiveness, and that continued influence arises only if a retraction is not believed. In two experiments, we found that source trustworthiness but not source expertise indeed influences retraction effectiveness, with retractions from low-trustworthiness sources entirely ineffective. We also found that retraction belief is indeed a predictor of continued reliance on misinformation, but that substantial continued influence effects can still occur with retractions designed to be and rated as highly credible.

Keywords Misinformation · Continued influence effect · Source credibility · Trustworthiness · Expertise · Persuasion

Introduction

Misinformation that is initially presented as true but is later revealed to be false is known to have an ongoing influence on inferential reasoning; this is known as the continued influence effect (CIE; Chan, Jones, Jamieson, & Albarracín, 2017; Johnson & Seifert, 1994; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Paynter et al., 2019; Walter & Murphy, 2018; Walter & Tukachinsky, 2020; Wilkes & Leatherbarrow, 1988). In the standard CIE paradigm, participants are presented with an event report (e.g., a report about a wildfire) that does or does not contain a critical piece of information, typically relating to the cause of the event (e.g., that the fire was intentionally lit). If the critical information is provided, it is or is not subsequently retracted. Participants' event-related reasoning is then probed via questionnaire (e.g., asking them whether someone deserves to be punished for the fire). Results typically show that a direct retraction significantly reduces reliance on the critical information relative to the no-retraction control condition, but does not eliminate the

influence down to the no-misinformation baseline (e.g., Ecker, Hogan, & Lewandowsky, 2017; Ecker, Lewandowsky, & Apai, 2011). Continued influence has also been demonstrated with real-world news (Lewandowsky, Stritzke, Oberauer, & Morales, 2005), common myths (Ferrero, Hardwicke, Konstantinidis, & Vellido, 2020; Sinclair, Stanley, & Seli, 2019; Swire, Ecker, & Lewandowsky, 2017), political misconceptions (Ecker & Ang, 2019; also see Ecker, Sze, & Andreotta, 2021; Nyhan & Reifler, 2010; Wood & Porter, 2019), with subtle and implicit misinformation (Ecker, Lewandowsky, Chang, & Pillai, 2014; Rich & Zaragoza, 2016), false allegations (Thorson, 2016; but see Ecker & Rodricks, 2020), and when the misinformation is presented initially as a negation that is later reinstated (Gordon, Ecker, & Lewandowsky, 2019).

Much theorizing has focused on the role of memory and memory-updating processes, arguing that the CIE arises from either selective retrieval of the misinformation (Ecker, Lewandowsky, Cheung, & Maybery, 2015; Ecker, Lewandowsky, Swire, & Chang, 2011; Gordon, Quadflieg, Brooks, Ecker, & Lewandowsky, 2019; Rapp, Hinze, Kohlhepp, & Ryskin, 2014; Rich & Zaragoza, 2016; Swire et al., 2017; also see Ayers & Reder, 1998) or from integration failure when processing the retraction (Brydges, Gignac, & Ecker, 2018; Ecker et al., 2017; Gordon, Brooks, Quadflieg, Ecker, & Lewandowsky, 2017; Kendeou, Butterfuss, Kim, & van Boekel, 2019; Kendeou, Walsh, Smith, & O'Brien,

✉ Ullrich K. H. Ecker
ullrich.ecker@uwa.edu.au

¹ School of Psychological Science (M304), University of Western Australia, 35 Stirling Hwy, Perth 6009, Australia

2014). One factor that has not been adequately considered in this theorizing to date is the influence of believability, that is, both the credibility of the information sources and the plausibility of the presented information per se. Part of the reason for this might be that most of the early CIE studies used reasonably plausible retractions that came from arguably high-credibility sources. For example, the seminal Johnson and Seifert (1994) study found a complete null effect of a retraction delivered by a presumably credible source (a police officer investigating the circumstances of a fire) that should have made the retraction more believable than the original misinformation, which came from an unspecified source. If the CIE occurs reliably even with plausible retractions from credible sources, then theoretical models need to be able to explain the effect without affording a dominant role to believability factors. Yet, this does not mean that these factors are not influential – for example by modulating the cognitive processes involved in the CIE – and so the present study was designed to investigate the impact of source credibility and retraction belief on the CIE.

Regarding the impact of source credibility, the source of information is known to be a strong determinant of persuasion and belief formation generally (Briñol & Petty, 2009; Cone, Flaharty, & Ferguson, 2019; Kumkale, Albarracín, & Seignourel, 2010; Pornpitakpan, 2004), although credibility may not have strong effects if readers pay little attention to the source (Sparks & Rapp, 2011; van Boekel, Lassonde, O'Brien, & Kendeou, 2017), or when the sources are media channels (Dias, Pennycook, & Rand, 2020). In the post-event misinformation literature (see Loftus, 2005), it is known that the impact of post-event misinformation on people's event recollection is partially dependent on the credibility of the misinformation source (Dodd & Bradshaw, 1980; Echterhoff, Hirst, & Hussy, 2005; Underwood & Pezdek, 1998), although there is also evidence that source credibility plays only a limited role in the formation of false memories (Wagner & Skowronski, 2017).

In the context of the CIE, there is some preliminary evidence that source credibility also plays a role. Cook and Lewandowsky (2016) showed that low source credibility can drive seemingly irrational belief updating in that corrective communications can have counter-intentional effects. Some have argued that if misinformation comes from a source that is perceived to be much more credible than the source of the retraction, continued influence can in fact be rational (Connor Desai, Pilditch, & Madsen, 2020; also see Jern, Chang, & Kemp, 2014). A recent analysis by Walter and Tukachinsky (2020) suggested that credibility of the misinformation source is a strong determinant of later continued influence effects (also see Swire-Thompson, Ecker, Lewandowsky, & Berinsky, 2020), but rather unintuitively found that credibility of the retraction source had no significant impact on the size of the continued influence effect.

The most direct investigations of the potential link between source credibility and the CIE were conducted by Guillory and Geraci (2010, 2013). In their 2010 study, they asked participants to explain *why* they thought a retraction was being given. While the majority of participants believed the retraction was being given to correct an earlier mistake, a substantial subset indicated the retraction may be an intentional cover-up, and these participants were more likely to rely on the original misinformation in their reasoning. In their 2013 study, Guillory and Geraci assessed whether manipulating the credibility of the retraction source directly would influence subsequent reliance on misinformation. Participants were presented with a report containing the critical piece of information that a politician was seen accepting a bribe, which was then retracted by a source low versus high in credibility. In their Experiment 1, it was found that retractions were significantly more effective when they came from a high-credibility source, resulting in fewer references to the misinformation concerning the bribe. Their Experiments 2 and 3 tested whether two dimensions of source credibility – expertise and trustworthiness (Pornpitakpan, 2004) – had differential impact on retraction effectiveness, by manipulating one factor while maintaining the other at a moderate level. It was found that the greater effectiveness of high-credibility sources was driven solely by perceived trustworthiness, not perceived source expertise. In other words, trustworthiness influenced retraction effectiveness (with source expertise held constant), while expertise did not influence retraction effectiveness (with trustworthiness held constant).

Regarding plausibility of the presented information per se, Lewandowsky et al. (2005) investigated people's belief in misinformation relating to the 2003 Iraq invasion, and showed that continued influence was stronger in people who were less sceptical of the original misinformation, suggesting that misinformation plausibility is one factor determining continued influence. Hinze, Slaten, Horton, Jenkins, and Rapp (2014) demonstrated that participants extracted inaccurate information from a provided text and used it in a subsequent general-knowledge test only if the information was plausible. Regarding the believability of the retraction, Ecker, Lewandowsky, and Tang (2010) reported that there was no difference in misinformation reliance between participants who did versus those who did not believe the retraction – however, the sample size in that study was small. O'Rear and Radvansky (2020) recently provided some evidence that the CIE may arise exclusively when participants do not believe the retraction, a finding that undermines existing theoretical models of the CIE and suggests that much more weight should be put on believability factors.

The present study was designed to answer two main questions. The first relates to the findings of Guillory and Geraci (2013): Is it indeed the case that perceived trustworthiness of the retraction source determines the effectiveness of a

retraction and thus the size of the CIE, while perceived source expertise plays no role whatsoever? The second question relates to the findings of O’Rear and Radvansky (2020): Is it indeed the case that a lack of belief in the retraction is a prerequisite for a CIE to occur?

Experiment 1

As mentioned earlier, it has been proposed that source credibility comprises two dimensions: trustworthiness and expertise. Trustworthiness can be defined as the willingness of a source to provide information that they believe to be correct; expertise refers to the ability of the source to make accurate and reliable assertions, and derives from an accumulation of competences or knowledge through experience and education (Pornpitakpan, 2004; Stadler & Bromme, 2014; Sternthal, Phillips, & Dholakia, 1978). While there is evidence to suggest that increasing either dimension independently of the other will improve the acceptance and influence of a persuasive message (e.g., Wiener & Mowen, 1986), there has historically been debate over whether or not the trustworthiness and expertise components of source credibility have differential weights. For instance, McGuire (1969) argued that expertise was more important than trustworthiness when it comes to persuasion; in line with this, Hovland and Mandell (1952) found that the influence of trustworthiness was negligible in comparison to expertise. By contrast, Cantor, Alfonso, and Zillmann (1976) reported no impact of source expertise on persuasion, and Hovland and Weiss (1951) reported strong effects of trustworthiness. McGinnies and Ward (1980) found that although expertise was influential, trustworthy sources were highly persuasive even when expertise was low. In her comprehensive review, Pornpitakpan (2004) concluded that both credibility dimensions are generally associated with persuasive effects, but that future research should aim to better separate them.

Assuming that expertise generally does influence persuasion, Guillory and Geraci’s (2013) Experiment 2 findings of a null effect of expertise were arguably somewhat surprising. In fact, recommendations to seek out fact-checks from experts are common, and common sense would suggest that advice from experts on complex matters should be more influential than advice from non-experts, especially when it comes to fact-checks (Kim, Moravec, & Dennis, 2019; but also see Lewandowsky, Ecker, & Cook, 2017). One possible explanation for the null effect of expertise in Guillory and Geraci’s Experiment 2 may lie in the way expertise was operationalized. As mentioned earlier, the misinformation in that study concerned a politician taking a bribe. An example of one of the used sources rated low on expertise (with moderate trustworthiness) was the wife of the politician’s political rival, whereas a high-expertise (moderate-trustworthiness) source

was the accused politician’s election campaign manager or even the politician himself. While these sources do presumably differ in expertise regarding law and the political process, it is not clear that this would directly impact their ability to provide correct information concerning the alleged bribery. Instead, a more important factor determining their ability to provide correct information is whether or not they were present to witness the event where the bribe was thought to occur. In other words, expertise in this study was operationalized as “involvement in the events.” This deviates from the common interpretation of expertise, namely skills and knowledge acquired through experience and education (Pornpitakpan, 2004), applied in other studies. For example, in Wiener and Mowen (1986), where the key information concerned the mechanical condition and value of an automobile, the expertise manipulation concerned the level of training and certification of the mechanic who was assessing the vehicle (also see McGinnies & Ward, 1980). Moreover, despite the fact that high- and low-expertise sources in Guillory and Geraci’s Experiment 2 were rated as neutral in terms of trustworthiness, arguably the politician, his campaign manager, and the wife of the politician’s rival would all have conflicts of interest that make trustworthiness questionable. It is therefore possible that a manipulation of expertise more in line with its common interpretation, and avoiding any confounding influence of trustworthiness, would yield a significant influence of source expertise on retraction effectiveness.

Experiment 1 therefore aimed to conceptually replicate the findings of Guillory and Geraci (2013) using a clearer manipulation of both expertise and trustworthiness in a fully crossed design that also allowed us to test for an interaction. Furthermore, we used a range of scenarios, each featuring a piece of critical information that was or was not subsequently retracted using scenario-specific retraction sources, covering a range of situations resembling those seen in actual news articles. It is possible that the nature of the scenario influences the degree to which expertise or trustworthiness is influential, so using multiple scenarios served to minimise scenario-specific effects and maximise generalizability. Lastly, we also obtained a direct measure of belief in the critical information and belief in the retraction, to explore whether reliance on misinformation is influenced by the degree of belief in the retraction (as suggested by O’Rear & Radvansky, 2020).

Method

The experiment used a within-subjects design with the factors source expertise (low vs. high) and source trustworthiness (low vs. high), in addition to two control conditions. One control condition was a no-retraction condition; the other control condition was designed to represent the best possible retraction from a source very high in both expertise and trustworthiness (the reason this was not used as the high-expertise/

high-trustworthiness condition of the regular 2×2 design was to preserve the symmetry of that design, that is, to ensure that variations in one factor did not entail variations in the other, based on pilot ratings of the sources; see *Results* for details). Participants were presented with one scenario for each of the six conditions. The main dependent variable – participants' reliance on the critical information – was measured using a questionnaire assessing participants' inferential reasoning regarding each scenario. Additionally, participants' belief in the critical information and their belief in the retraction were assessed directly.

Pilot study

Prior to running the main experiment, a pilot study was conducted in order to select the scenarios and retraction sources. Participants rated retraction sources on perceived expertise and trustworthiness within a particular scenario context. The aim was to select six scenarios with at least four different sources that varied sufficiently on both dimensions, from a pool of 14 scenarios with an average of nine retraction sources each.

Participants The online crowd-sourcing platform CrowdFlower was used to recruit participants located in Australia, the USA, and the UK. 100 participants completed the pilot survey in exchange for US\$2. The data were screened for signs of negligent responding, as indicated by completion times significantly faster than the average or by low response variability. Mean completion time was $M = 22$ min ($SD = 11.6$). The average response variability (the standard deviation across 240 total responses per participant, with responses on a five-point scale) was $M = 1.05$ ($SD = .29$). Participants more than one standard deviation below the mean on either of these criteria (i.e., completion < 10 min; response variability < 0.76), were excluded from analysis ($n = 20$). This left a sample of $N = 80$, comprising 22 male and 58 female participants, ranging in age from 22 to 69 years ($M = 43.05$, $SD = 11.92$).

Materials and procedure The pilot study was run as an online survey using Qualtrics software (Qualtrics, Provo, UT, USA). Participants were presented with 14 brief scenarios concerning an unfolding news event. Each scenario consisted of eight sentences. The critical information was embedded in the second sentence (e.g., *The Charleston Water Department has been forced to shut down water intake from the Elk River following reports of large-scale fish deaths in the waterway. It is believed that the fish deaths were caused by contamination with industrial pollutants from a nearby mining company*). The retraction was presented in the second-last sentence; it came from a named but otherwise unspecified source (e.g., *Todd Hunter, [source], explained that "there was no contamination from mining operations"*). Following each scenario, participants were given a list of up to ten

potential sources for the retraction – for example, suggestions regarding the identity of "Todd Hunter" included a "Water Department biologist" or a "mine employee." Participants rated the expertise and trustworthiness of the potential sources on a five-point scale from 0 (very low) to 4 (very high). The order of presentation for both the scenarios and sources was randomized.

Main study

Participants An a priori power analysis (using G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007) suggested that to detect an effect of size $\eta_p^2 = .145$ – the average size of the credibility effects in Guillory and Geraci's (2013) Experiments 1 and 3 – would require a minimum sample size of 50 participants (a priori repeated-measures ANOVA; $\alpha = .05$; $1 - \beta = .8$; one group; two measurements; no non-sphericity correction). A total of $N = 53$ undergraduate students from the University of Western Australia were recruited for participation in return for course credit. The sample consisted of 20 male and 33 female participants, with an average age of $M = 18.64$ years ($SD = 2.31$; range 17–32).

Materials Six scenarios and the sources for the five retraction conditions were selected based on the pilot-study ratings. Sources were chosen to maximize the expertise and trustworthiness differences, while ensuring that, on average, there were no substantial variations in one dimension across variations of the other. The pilot rating data are presented in the *Results* section. The six selected scenarios existed in both retraction and no-retraction versions. Two additional scenarios existed only in a no-retraction version; these had only seven sentences and were merely used as fillers to avoid the build-up of strong retraction expectations. All scenarios are provided in the Online Supplementary Material (OSM) at <https://osf.io/awsf5>. The assignment of scenarios to conditions, and the presentation order of scenarios/conditions were counterbalanced across participants; to this end, participants were randomly allocated to one of six survey versions, also detailed in the OSM.

Questionnaires There were three test questionnaires. The first questionnaire featured a recall question and several inferential reasoning questions relating to each scenario. The recall question simply required the participant to summarize the scenario from memory. The inference questions included four open-ended questions, designed to elicit responses related to the critical information, while also allowing for unrelated responses (e.g., *What could be done to prevent such incidents in the future?*), and three rating scales, requiring participants to indicate their level of agreement with a statement on a scale from 0 to 10 (e.g., *Should the mining company be fined for the incident?*). The questionnaire can be found in the OSM.

The second questionnaire re-presented the scenarios involving a retraction (from the specified source) in a slightly abbreviated form, and asked the participant to indicate their belief in the initial claim (i.e., the critical information) and their belief in the retraction on a scale from 0 (not at all) to 10 (very strong). The first two questionnaires were given in separate booklets following the order of scenarios in the study phase.

The final questionnaire was a manipulation check and delivered electronically via Qualtrics. Participants were asked to rate the expertise and trustworthiness of each of the five retraction sources for each scenario, replicating the pilot study. The order in which the scenarios and sources were presented was randomised.

Procedure Participants read an ethics-approved information sheet and provided written consent. After being randomly assigned to one of the six survey versions, participants were presented with the eight scenarios on a computer screen in an individual testing booth. Scenarios were presented sentence-by-sentence; each sentence was shown for a set duration (calculated as 400 ms per word). Participants were then presented with an unrelated distractor task for 10 min. Finally, participants completed the three questionnaires. The entire experiment took approximately 1 h to complete.

Results

Pilot study

The expertise and trustworthiness ratings of the five retraction conditions, averaged across the six scenarios selected for inclusion in the main experiment, are given in the left panel of Table 1.

Table 1 Descriptive statistics from pilot and main studies for retraction-source conditions averaged across the six scenarios selected for inclusion in the main study

	Pilot study				Main study			
	Expertise		Trustworthiness		Expertise		Trustworthiness	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
LELT	1.67	0.30	1.19	0.43	1.62	0.54	1.32	0.57
LEHT	1.80	0.24	2.13	0.21	1.53	0.48	1.92	0.64
HELT	2.72	0.22	1.27	0.34	2.64	0.58	1.20	0.78
HEHT	2.76	0.35	2.44	0.26	2.69	0.43	2.35	0.46
HEHT+	3.33	0.29	3.04	0.36	3.47	0.45	3.02	0.58

LELT low expertise, low trustworthiness; LEHT low expertise, high trustworthiness; HELT high expertise, low trustworthiness; HEHT high expertise, high trustworthiness; HEHT+ highest expertise and trustworthiness (“best retraction” control condition)

The scale was 0–4

Main study

Source ratings (manipulation check) The first step of the analysis was to confirm that the ratings of source expertise and trustworthiness replicated the pilot study. Mean ratings across conditions are given in the right panel of Table 1. By and large, sources were perceived very similarly in the pilot and main studies. The five conditions differed significantly in terms of expertise, $F(4,208) = 217.76$, $MSE = .16$, $p < .001$, $\eta_p^2 = .81$; a planned contrast analysis showed that both low-expertise conditions differed from all high-expertise conditions, all $F_s(1,52) \geq 119.79$, all $p_s < .001$. The five conditions also differed significantly in terms of trustworthiness, $F(4,208) = 101.08$, $MSE = .30$, $p < .001$, $\eta_p^2 = .66$; a planned contrast analysis showed that both low-trustworthiness conditions differed from all high-trustworthiness conditions, all $F_s(1,52) \geq 25.93$, all $p_s < .001$.

Questionnaire coding The open-ended recall and inference questions were coded by two scorers blind to the experimental conditions, following a standardized guide. All scoring discrepancies were resolved through discussion. Any unambiguous reference to the critical information was scored 1 (e.g., *The mining company is to blame for the pollution in the river*). References to the critical information suggesting an ambiguous level of endorsement were scored 0.5 (e.g., *The mining company may have been involved?*). Responses were scored 0 where the critical information was not mentioned or was specifically disavowed (e.g., *The mining company was at first suggested to be involved but they were proven not to be*). Responses to the open-ended recall and inference questions and rating-scale responses were combined and transformed into an inference score on a continuous 0–10 scale, with higher scores indicating a higher level of endorsement of the critical information.

Belief ratings The mean direct belief scores for the critical information and the retraction across conditions are presented in Fig. 1. A two-way repeated-measures ANOVA with factors information (critical information; retraction) and condition (LELT, LEHT, HELT, HEHT, HEHT+) showed a main effect of information, indicating that belief in the critical information was significantly greater than belief in the retraction across conditions, $F(1,52) = 56.63$, $MSE = 9.21$, $p < .001$, $\eta_p^2 = .52$. However, there was no main effect of condition, $F(4,208) = 1.83$, $MSE = 2.31$, $p = .124$, and no interaction, $F(4,208) = 1.05$, $MSE = 6.94$, $p = .381$.

Inference scores Inference scores were the main dependent variable; mean inference scores (and 95% confidence intervals [CIs]) across conditions were $M_{\text{NoR}} = 5.31$ [4.79–5.82]; $M_{\text{LELT}} = 5.04$ [4.53–5.55]; $M_{\text{LEHT}} = 4.37$ [3.76–4.98]; $M_{\text{HELT}} = 5.19$ [4.71–5.68]; $M_{\text{HEHT}} = 4.33$ [3.82–4.85]; $M_{\text{HEHT+}} = 4.55$ [4.03–5.07]; they are shown graphically in

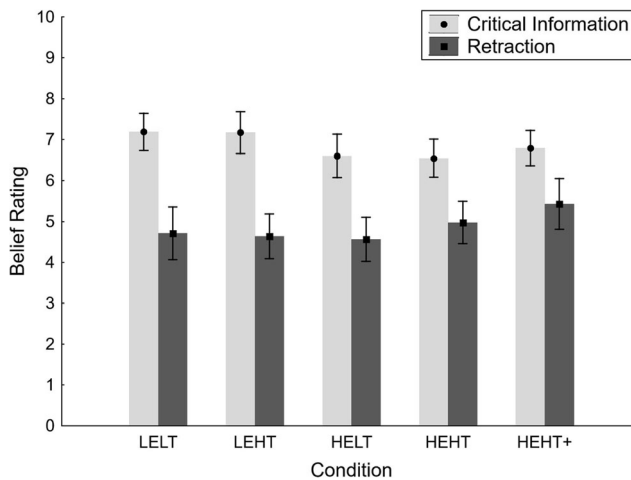


Fig. 1 Mean belief ratings across conditions in Experiment 1. *LELT* low expertise, low trustworthiness; *LEHT* low expertise, high trustworthiness; *HELT* high expertise, low trustworthiness; *HEHT* high expertise, high trustworthiness; *HEHT+* highest expertise and trustworthiness. Error bars indicate 95% confidence intervals

Fig. 2. Focusing on the retraction conditions of the core 2 (source expertise: low, high) \times 2 (source trustworthiness: low, high) design, a two-way repeated-measures ANOVA yielded a main effect of source trustworthiness: participants made significantly fewer references to the critical information when the retraction came from high-trustworthiness sources rather than low-trustworthiness sources, $F(1,52) = 8.25$, $MSE = 3.81$, $p = .006$, $\eta_p^2 = .14$. Expertise, however, did not have a significant effect, nor was there an interaction effect, $F_s < 1$.

To compare retraction conditions against control, we also ran a one-way repeated-measures ANOVA featuring all conditions; the main effect of condition was marginal, $F(5,255) = 2.25$, $MSE = 4.32$, $p = .050$, $\eta_p^2 = .04$. Planned comparisons between the pooled high- and low-trustworthiness conditions

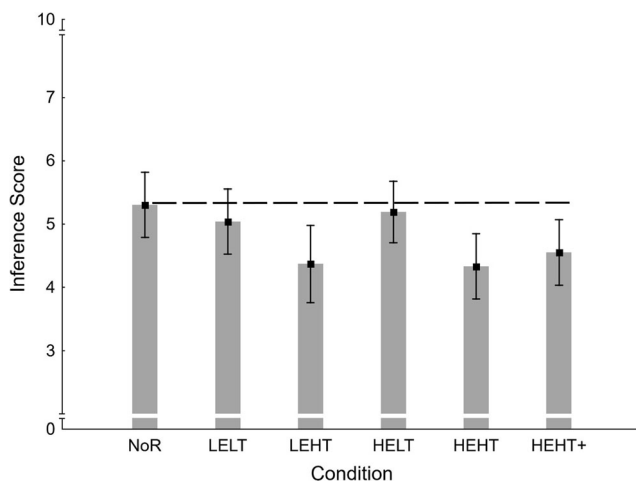


Fig. 2 Mean inference scores across conditions in Experiment 1. *noR* no-retraction control; *LELT* low expertise, low trustworthiness; *LEHT* low expertise, high trustworthiness; *HELT* high expertise, low trustworthiness; *HEHT* high expertise, high trustworthiness; *HEHT+* highest expertise and trustworthiness. The horizontal line reflects the control condition baseline. Error bars indicate 95% confidence intervals

and the control condition showed that the high-trustworthiness conditions differed significantly from control, $F(1,51) = 7.16$, $p = .010$, whereas the low-trustworthiness conditions did not, $F < 1$.¹ By contrast, the high-expertise conditions did not differ significantly from control, and neither did the low-expertise conditions, $F_s(1,51) \leq 3.50$, $p_s \geq .067$. This implies that only retractions from high-trustworthiness sources produced a significant reduction in misinformation reliance.

Even though retraction belief did not vary with condition as expected, to test the relation between inferential-reasoning scores and belief in the critical information and belief in the retraction, we applied linear mixed-effects modelling, using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015). We specified survey version and participant ID (nested in survey version) as random effects, and the ratings of critical-information belief and retraction belief as fixed effects, predicting inference scores. We found that belief in the critical information did not significantly predict inference scores, $\beta = .06$, $SE = .06$, $t < 1$, but that retraction belief did, $\beta = -.28$, $SE = .05$, $t(221) = -5.33$, $p < .001$.

To answer the question of whether there was still a significant CIE after a strongly believed retraction, we inspected inference scores in the *LEHT*, *HEHT*, and *HEHT+* conditions associated with retraction-belief scores ≥ 8 . Mean inference scores (and 95% CIs) were $M_{LEHT} = 3.81$ ([2.32–5.30]; $n = 8$); $M_{HEHT} = 2.86$ ([1.56–4.16]; $n = 10$), and $M_{HEHT+} = 3.18$ ([2.21–4.14]; $n = 15$). Unfortunately, Experiment 1 did not include a no-misinformation control condition, so it is difficult to gauge where the baseline would be, but it cannot be ruled out that these restricted samples did not show a significant CIE.

Discussion

Experiment 1 demonstrated that perceived source trustworthiness has greater impact on retraction effectiveness than perceived source expertise: Only retractions from trustworthy sources reduced reliance on misinformation, whereas source expertise played no discernible role. This conceptually replicates Guillory and Geraci (2013), despite the different operationalizations of expertise in the two studies, and is also in line with a large body of research from the broader persuasion literature (see Pornpitakpan, 2004). The most likely reason for the ineffectiveness of retractions from low-trustworthiness sources is arguably enhanced scepticism – it is known that scepticism towards misinformation reduces post-retraction misinformation impact (Fein, McCloskey, & Tomlinson, 1997; Lewandowsky et al., 2005; Schul, Mayo, & Burnstein, 2008), and it is logical to assume the same applies to the impact of the retraction.

¹ Contrasts of individual high-trustworthiness conditions versus control yielded *LEHT*: $F(1,51) = 4.17$, $p = .046$; *HEHT*: $F(1,51) = 5.30$, $p = .025$; *HEHT+*: $F(1,51) = 4.32$, $p = .043$.

The null effect of expertise may have arisen because of participants not paying much attention to this dimension. Although Braasch, Rouet, Vibert, and Britt (2012) suggested that text discrepancies, such as retractions, promote attention to source information, the lack-of-attention claim is in line with van Boekel et al. (2017), who found source-credibility effects on the processing of refutation processing only when participants were explicitly instructed to pay attention to source credibility (also see Sparks & Rapp, 2011, for a similar finding with persuasive text comprehension). Even if people pay some attention to source expertise, they may have a tendency to assign little weight to this dimension because seemingly objective expertise cues can be misleading (e.g., when vested interest groups present “fake experts”; Cook, Lewandowsky, & Ecker, 2017; Diethelm & McKee, 2009) and cues of subjective expertise are often of little value given that self-proclaimed experts often lack knowledge while being overly confident in their expertise (Kuklinski, Quirk, Jerit, Schwieder, & Rich, 2000; Motta, Callaghan, & Sylvester, 2018).² We also know that if people question the validity of a retraction, they do so primarily due to a concern that the correction may be a deceitful cover-up rather than unintentionally inaccurate (Guillory & Geraci, 2010) – perhaps because they perceive deceit as more likely or its implications as more severe – and hence source trustworthiness but not expertise moves into the focus of attention.

In general, the effect of the retractions in Experiment 1 was relatively small. While it is difficult to say what the baseline was, due to the lack of a no-misinformation control condition, comparable studies have yielded greater retraction effects, with retractions approximately halving reliance on misinformation (Ecker, Lewandowsky, & Apai, 2011; Ecker et al., 2017; Rich & Zaragoza, 2016). One reason for this discrepancy may be that the retractions in the present study provided no context and no supportive arguments, and it is known that such terse retractions are relatively ineffective (Ecker, O’Reilly, Reid, & Chang, 2020; Swire et al., 2017). In line with this, participants indicated that they believed the critical information more than the retraction across all conditions. Moreover, the mixed-effects modelling demonstrated that retraction belief was an important driver of post-retraction misinformation reliance, and the misinformation inference scores were particularly low when credible retractions were strongly believed. Even though CIEs were generally observed even with retractions from trustworthy sources, this pattern is in line with suggestions by O’Rear and Radvansky (2020) that CIEs do not arise if participants fully believe the retraction.

² Additionally, objective expertise cues may evoke trust in some but distrust in others (Fischer, 2019; Gauchat, 2012; Hamilton, Hartter, & Saito, 2015); this relates to real-world CIEs in the context of science denial, where people who distrust science are most likely to dismiss scientific rebuttals of false claims and misconceptions (e.g., Cook & Lewandowsky, 2016).

Overall, we believe that when seen in conjunction with previous research – and in particular Guillory and Geraci (2013) – the findings of Experiment 1 allow the conclusion that perceived source trustworthiness generally matters for corrections, whereas perceived source expertise does not. We nevertheless decided to run a second experiment. However, the primary aim of Experiment 2 was not so much to obtain another replication of trustworthiness effects (or the absence of expertise effects), but instead – inspired by the findings of O’Rear and Radvansky (2020) – focussed more on our second research question, namely, whether continued influence arises with highly credible retractions. At the same time, Experiment 2 attempted to address a methodological issue shared by both Experiment 1 and O’Rear and Radvansky’s study.

Experiment 2

The methodological issue with Experiment 1 (and also O’Rear & Radvansky, 2020) was that critical-information and retraction beliefs were measured immediately after inferential reasoning. This means that participants may have rated their beliefs to retrospectively match their reasoning – for example, if a participant’s responses to the inferential reasoning questions reflected reliance on the retracted misinformation, they may have indicated strong critical-information belief and/or low retraction belief in order to justify their prior responses. This methodological problem is difficult to avoid entirely; however, it may be preferable to obtain online measures of credibility during encoding rather than retrospective measures. Thus, in Experiment 2, we obtained credibility ratings for every single message presented during the study phase. While, of course, one could argue that in this case later reasoning responses could be adjusted to match prior credibility ratings, we believe that risk to be much smaller, given the longer interval between the two ratings and the fact that *all* messages were rated for credibility. Obtaining online credibility ratings also allowed us to test whether continued influence would be observed with retractions from highly credible sources that attracted high credibility ratings and can thus be assumed to have been believed. According to O’Rear and Radvansky (2020), there should be no continued influence – that is, a significant difference between a retraction and a no-misinformation condition – in such cases.

Method

Using four of the scenarios from Experiment 1, Experiment 2 set out to test the effects of source trustworthiness and the credibility of critical-information and retraction messages. To this end, Experiment 2 used four conditions in a within-subjects design: a no-misinformation control condition, a no-

retraction control condition, and two retraction conditions using retraction sources with low versus high trustworthiness (but high expertise, thus replicating the HELT and HEHT+ conditions of Experiment 1). Inferential reasoning was measured using the same questionnaire as in Experiment 1.

Participants A total of $N = 68$ participants were recruited for Experiment 2; participants were undergraduate students from the University of Western Australia who had not participated in Experiment 1. The sample comprised 29 males and 39 females; mean age was $M = 21.99$ years ($SD = 7.50$; range: 18–57). Unfortunately, due to technical error, credibility responses were not recorded for three participants; these were omitted from analyses involving credibility.

Materials Four scenarios from Experiment 1 were selected³ and condensed into five messages each, such that messages 1 and 3 contained multiple sentences (previously presented individually in Experiment 1). The critical information was still presented in the second message, and the retraction was still presented in the second-last message; these messages were simply omitted in the no-misinformation and no-retraction conditions. The assignment of scenarios to conditions, as well as the presentation order of scenarios/ conditions, were counterbalanced across participants. To this end, participants were randomly allocated to one of four survey versions, detailed in the OSM.

Questionnaires Questionnaires to measure inferential reasoning were identical to the ones used in Experiment 1.

Procedure The procedure was identical to Experiment 1 with the following exceptions: Participants only read four scenarios and only completed one test questionnaire per scenario. Scenarios were presented message-by-message on a screen using OpenSesame software (Mathôt, Schreij, & Theeuwes, 2012), and participants rated the credibility of each message on a 0 (non-credible) to 3 (fully credible) scale,⁴ using the number keys on a standard QWERTY keyboard number pad. The experiment took approximately 20 min to complete.

Results

Credibility ratings

Mean credibility ratings for the critical information were $M = 2.22$ ($SD = 0.82$) for the no-retraction condition, $M = 2.06$ ($SD = 0.92$) for the low-trustworthiness condition, and $M = 2.22$

($SD = 0.84$) for the high-trustworthiness condition. Retraction-credibility ratings were $M = 1.72$ ($SD = 1.05$) for the low-trustworthiness condition and $M = 1.92$ ($SD = 1.04$) for the high-trustworthiness condition. A one-way ANOVA across all five mean credibility ratings yielded a main effect of condition, $F(4,256) = 3.74$, $MSE = 0.76$, $p = .006$, $\eta_p^2 = .06$. Similar to Experiment 1, the critical information was rated as more credible than the retraction across conditions, $F(1,64) = 7.27$, $MSE = 1.25$, $p = .009$. The difference between low- and high-credibility retractions was non-significant, $F(1,64) = 1.43$, $MSE = 0.91$, $p = .236$.

Inference scores

Inference scores were again the main dependent variable; mean inference scores across conditions were $M_{NoMI} = 3.13$ [2.75–3.52]; $M_{NoR} = 4.93$ [4.39–5.47]; $M_{LT} = 5.88$ [5.33–6.42]; $M_{HT} = 5.24$ [4.74–5.73]; they are shown graphically in Fig. 3. A one-way repeated-measures ANOVA featuring all conditions returned a significant main effect of condition, $F(3,201) = 16.79$, $MSE = 5.58$, $p < .001$, $\eta_p^2 = .20$. Planned comparisons showed that, as expected, the no-misinformation condition differed from all three conditions that presented the critical information, all $F_s(1,67) \geq 31.33$, $p_s < .001$. The no-retraction condition differed marginally from the low-trustworthiness condition (in the opposite direction than expected), $F(1,67) = 4.00$, $p \leq .050$, while there was no difference between no-retraction and high-trustworthiness conditions, $F < 1$, indicating that the retractions were entirely ineffective. The retraction conditions did not differ from each other, $F(1,67) = 2.65$, $p = .108$.

To test the relation between inferential-reasoning scores and belief in the critical information and belief in the retraction, we again applied linear mixed-effects modelling. We

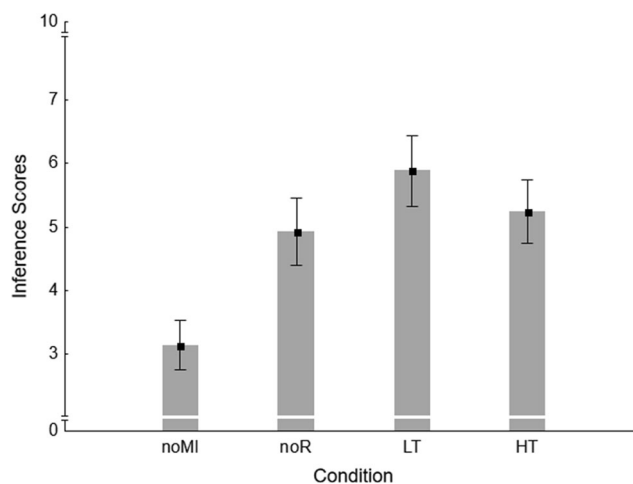


Fig. 3 Mean inference scores across conditions in Experiment 2. *noMI* no-misinformation control; *noR* no-retraction control; *LT* low trustworthiness; *HT* high trustworthiness. Error bars indicate 95% confidence intervals

³ Scenarios were chosen in an ad hoc fashion, without looking at scenario-specific data from Experiment 1.

⁴ We used a different scale from Experiment 1 (which used a 0–4 scale) to avoid offering a mid-point, in order to encourage more thorough engagement.

specified survey version and participant ID (nested in survey version) as random effects, and the ratings of critical-information credibility and retraction credibility as fixed effects, predicting inference scores. Unlike the analysis in Experiment 1, this analysis included no-misinformation and no-retraction conditions; we modelled these conditions by setting credibility to zero for non-presented misinformation and/or retractions. Unlike Experiment 1, we found that the credibility of the critical information predicted inference scores, $\beta = .06$, $SE = .01$, $t(260) = 6.38$, $p < .001$, but that the credibility of the retraction did not, $\beta = .01$, $SE = .01$, $t < 1$. That being said, restricting the analysis to retraction conditions (in line with Experiment 1), credibility of the critical information was no longer a significant predictor, $\beta = .02$, $SE = .02$, $t < 1$, whereas the credibility of the retraction became significant, $\beta = -.04$, $SE = .02$, $t(124) = -2.14$, $p = .034$, mirroring the results from Experiment 1.

To answer the question of whether there was still a significant CIE after a strongly believed retraction, we inspected inference scores in the high-trustworthiness condition associated with the maximum retraction-credibility rating of 3. The mean inference score (and 95% CI) was $M_{HT} = 4.83$ ([3.77–5.89]; $n = 23$), which lies substantially above the no-misinformation baseline of $M_{NoMI} = 3.13$ [2.75–3.52]. A significant CIE was thus present even with a fully credible retraction from a trustworthy source.

Discussion

Experiment 2 aimed to test whether retraction credibility determined the extent to which participants continued to rely on retracted misinformation. Retractions from sources high versus low in trustworthiness were equally ineffective; thus Experiment 2 did not replicate the main finding of Experiment 1 that retractions from trustworthy sources reduced reliance on misinformation. Also, unlike Experiment 1 – and unlike much other previous research – retractions were found not to reduce reliance on misinformation at all. This is not entirely unprecedented, as the seminal study by Johnson and Seifert (1994) also reported entirely ineffective retractions. In addition to the reasons already mentioned in the Discussion of Experiment 1, namely use of retractions that provided no context and no supportive arguments, we speculate that the online credibility ratings employed in Experiment 2 served to further enhance participants' scepticism during encoding, leading them to by-and-large dismiss the retractions (see Fein et al., 1997; Lewandowsky et al., 2005; Schul et al., 2008). It seems especially likely that the online credibility ratings in Experiment 2 created unusual processing conditions given a finding by Ithisuphalap, Rich, and Zaragoza (2020), who reported that assessing one's belief in an event cause (after encoding an event report but before receiving a correction) enhanced the effectiveness of the subsequent correction.

Ithisuphalap et al. argued that scrutinizing one's belief leads to a more elaborate representation that promotes conflict detection when the correction is provided, facilitating belief revision (also see Ecker et al., 2017; Kendeou et al., 2014, 2019). This clearly differs from what we observed in Experiment 2, which suggests that our online ratings promoted general scepticism that disproportionately impacted on correction endorsement.

In line with this, participants again indicated that they believed the critical information more than the retraction across conditions. The mixed-effects modelling demonstrated that while the credibility of the critical information was a strong determinant of references to the critical information when considering all conditions, when restricted to the retraction conditions, only retraction credibility had a significant, albeit small, influence on post-retraction misinformation reliance, in line with Experiment 1. However, even in cases where a retraction from a trustworthy source was rated as fully credible, a continued influence effect was observed. This conflicts with O'Rear and Radvansky's (2020) suggestion that CIEs do not arise if participants fully believe a retraction.

General discussion

The first conclusion to be drawn from this study, when seen in conjunction with previous research (in particular, Guillory & Geraci, 2013), is that the trustworthiness of retraction sources matters, at least under standard processing conditions such as those employed in Experiment 1. In Experiment 1, we found that only retractions from sources perceived to be trustworthy reduced reliance on retracted misinformation, while the perceived source expertise had no impact. As such, retractions from expert sources were ineffective if trustworthiness was low. This pattern replicates Guillory and Geraci (2013). Theoretically, this finding may indicate that trustworthiness is a primary dimension of source credibility, as suggested by McGinnies and Ward (1980) – that is, it may be impossible for credibility to be high if perceived trustworthiness is low, whereas credibility can be high if perceived expertise is low, based on high levels of trustworthiness alone. This is potentially concerning, because rebuttal messages from non-expert sources may have undue impact to the extent that the source is perceived as trustworthy – consider celebrities who publicly endorse and promote conspiracy theories (e.g., see Cockerell, 2020, for current examples regarding the 5G COVID-19 conspiracy theory) or “trusted locals” used by media outlets to oppose a scientific consensus (such as a regular swimmer at Australia's iconic Bondi Beach claiming they had seen no evidence of rising sea levels; Vasek & Franklin, 2010). With regards to Experiment 2, the fact that we did not observe a significant trustworthiness effect should in our view not be over-rated, as retractions were generally ineffective in that

experiment, presumably driven by the online credibility ratings inducing high levels of scepticism, which in turn may have led to immediate dismissal of retractions. However, this remains speculative and additional research is needed to draw stronger conclusions.

An effect of perceived trustworthiness meshes well with the evidence provided by both experiments that retraction belief is an important determinant of post-retraction misinformation reliance and thus the CIE. Our results are consistent with the suggestion that people may continue to rely on retracted misinformation partly because they do not believe the retraction (see Guillory & Geraci, 2010; O’Rear & Radvansky, 2020): Across all conditions, belief in the retraction was significantly lower than belief in the original erroneous information, suggesting people considered the misinformation more likely to be true than the correction. It is not entirely clear why this effect was observed, especially as the misinformation was not linked to any source, meaning that readers seem to have simply inferred that the source was credible; the differential evaluation may have resulted from the brevity of the retractions, or from the fact that a retraction – unlike the initial critical information – is generally processed in light of its contradiction with the earlier information and is thus subject to immediate scrutiny. Thus, a retraction may by its very nature struggle with a “credibility handicap” as it by definition opposes previously provided information (see Lewandowsky et al., 2012; Seifert, 2002; Smithson, 1999). This handicap may have been enhanced by the scepticism induced by the online credibility ratings in Experiment 2, which could explain the ineffectiveness of retractions in that study. One way to further illuminate this in future research would be to manipulate both misinformation and correction sources in the same study.

Our results suggest that theoretical models of the CIE should place more emphasis on the role of retraction credibility. Future research should further scrutinize how credibility affects continued influence, with one obvious possibility being that people simply encode credible information from trustworthy sources more strongly – which means that source and message credibility may primarily affect ongoing reliance on misinformation by modulating the memory and integration processes that are assumed to underlie the CIE (Ecker, Lewandowsky, Swire, et al., 2011; van Boekel et al., 2017). Moreover, contemporary models of the CIE should acknowledge more explicitly that there can be instances where a CIE is rational – namely if a retraction is substantially less credible than the misinformation itself (see Connor Desai et al., 2020) – and that there may also be instances where no CIE should be expected or observed, namely if a convincing retraction comes from a trustworthy source and is strongly believed (see O’Rear & Radvansky, 2020).

That being said, however, we did observe continued influence under such conditions in Experiment 2. We acknowledge

that the online credibility rating task – implemented to disentangle measures of retrospective retraction belief and misinformation reliance used in previous work – may have created unusual encoding conditions. However, the fact that we obtained significant continued influence effects even with strongly believed retractions from highly credible sources speaks against a strong interpretation of O’Rear and Radvansky’s (2020) proposal that continued influence arises only if a retraction is not believed. Examining the relation between the CIE and retraction credibility must therefore remain a focus of future research. Clearly, perceived source credibility is not the only factor affecting continued influence. For example, in Rich and Zaragoza (2016), participants’ correction belief was not dependent on whether the corrected misinformation was explicitly provided or merely hinted at, even though continued influence was greater in the latter, implied-misinformation condition (presumably because a correction can more easily “target” an explicitly provided piece of misinformation).

Regarding the role of expertise, we note that the absence of evidence for a significant expertise effect of course does not constitute strong evidence for the absence of expertise effects in general. In particular, it remains a possibility that expertise may play a role when it comes to corrections communicated by organizations rather than individuals: Vraga and Bode (2017) as well as van der Meer and Jin (2020) found that refutations from health agencies such as the Centers for Disease Control and Prevention (CDC) were more effective than refutations from an individual. However, it is unclear whether these benefits were indeed caused by perceived expertise rather than trustworthiness and/or the status of an agency as opposed to an individual. The fact that van der Meer and Jin found refutations from a news agency (Reuters) to be as effective as refutations from the CDC suggests that expertise per se was not the primary factor. Moreover, the evidence that source credibility effects may occur only if information recipients actively monitor source credibility (e.g., Sparks & Rapp, 2011; van Boekel et al., 2017) suggests that perceived expertise might influence retraction effectiveness only in cases where expertise is made particularly salient.

The practical implications of this research are twofold: First, the observed effect of perceived trustworthiness suggests that efforts to reduce continued influence of misinformation should always take into account the trustworthiness of the retraction source. For example, public information campaigns seeking to counter misconceptions should consider using a variety of trusted messengers for different audiences (Moser, 2010; also see Krishna, 2018). While sources with high credibility on both dimensions (e.g., trusted health professionals or organizations) may be ideal choices even if expertise itself does not seem impactful (van der Meer & Jin, 2020; Vraga & Bode, 2017; also see Durantini, Albarracín, Mitchell, Earl, & Gillette, 2006), high-expertise sources should not be used

unless their perceived trustworthiness in the target audience is assessed. Second, the overall low effectiveness of retractions in the present study highlights that corrections cannot simply rely on source-credibility clues, but rather need to be designed to be convincing and persuasive in their own right (Ecker et al., 2020; Paynter et al., 2019).

Finally, while we and others have investigated the effects of source credibility by trying to manipulate credibility factors directly, more research is needed on the role of credibility proxies such as peer approval. For example, it has been suggested that social-media consumers rely on content endorsements (e.g., number of “likes”) as credibility cues (Chung, 2017; Messing & Westwood, 2012; also see Jost, van der Linden, Panagopoulos, & Hardin, 2018). Thus, factors such as endorsement of both misinformation and refutation messages in a social-media environment remain important targets for future misinformation research.

In conclusion, we set out to answer two questions: (1) Does perceived trustworthiness of a retraction source but not perceived expertise influence retraction effectiveness? (2) Is low retraction belief a prerequisite for continued influence to occur (as suggested by O’Rear & Radvansky, 2020)? In light of these questions, our study makes the following contributions: (1) It adds to the evidence that perceived source trustworthiness, but not perceived expertise, influences the effectiveness of a retraction. This is a novel contribution because our operationalization of expertise was much closer to the common understanding of the term than previous research, and this was the first CIE study to manipulate these two dimensions of source credibility in a fully-crossed design. (2) The study provides proof of concept that significant continued influence can arise even when highly credible retractions are strongly believed. This is important because previous research has suggested that continued influence simply results from lack of retraction belief (O’Rear & Radvansky, 2020). We therefore conclude that theoretical models of continued influence should take into account source credibility factors, without, however, affording them undue significance.

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/awsf5>. The experiments were not preregistered.

Acknowledgements The research was supported by the Australian Research Council under grants DP160103596 and FT190100708, awarded to the first author. We thank Charles Hanich for research assistance.

References

Ayers, M., & Reder, L. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model.

- Psychonomic Bulletin & Review*, 5, 1–21. <https://doi.org/10.3758/BF03209454>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Braasch, J. L. G., Rouet, J., Vibert, N., & Britt, M. A. (2012). Readers’ use of source information in text comprehension. *Memory & Cognition*, 40, 450–465. <https://doi.org/10.3758/s13421-011-0160-6>
- Briñol, P., & Petty, R. E. (2009). Source factors in persuasion: A self-validation approach. *European Review of Social Psychology*, 20, 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>
- Cantor, J. R., Alfonso, H., & Zillmann, D. (1976). The persuasive effectiveness of the peer appeal and a communicator’s first-hand experience. *Communication Research*, 3, 293–310. <https://doi.org/10.1177/009365027600300304>
- Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28, 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Chung, M. (2017). Not just numbers: The role of social media metrics in online news evaluations. *Computers in Human Behavior*, 75, 949–957. <https://doi.org/10.1016/j.chb.2017.06.022>
- Cockerell, I. (2020). Meet the celebrities pushing 5G coronavirus conspiracies to millions of fans. <https://www.codastory.com/waronscience/celebrities-5g-conspiracies/>
- Cone, J., Flaharty, K., & Ferguson, M. J. (2019). Believability of evidence matters for correcting social impressions. *Proceedings of the National Academy of Sciences*, 116, 9802–9807. <https://doi.org/10.1073/pnas.1903222116>
- Connor Desai, S., Pilditch, T., & Madsen, J. K. (2020). The rational continued influence of misinformation. *Cognition*, 205, 104453. <https://doi.org/10.1016/j.cognition.2020.104453>
- Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science*, 8, 160–179. <https://doi.org/10.1111/tops.12186>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12, e0175799. <https://doi.org/10.1371/journal.pone.0175799>
- Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *The Harvard Kennedy School (HKS) Misinformation Review*, 1. <https://doi.org/10.37016/mr-2020-001>
- Diethelm, P., & McKee, M. (2009). Denialism: what is it and how should scientists respond? *European Journal of Public Health*, 19, 2–4. <https://doi.org/10.1093/eurpub/ckn139>
- Dodd, D. H., & Bradshaw, J. M. (1980). Leading questions and memory: Pragmatic constraints. *Journal of Verbal Learning and Verbal Behavior*, 19, 695–704. [https://doi.org/10.1016/S0022-5371\(80\)90379-5](https://doi.org/10.1016/S0022-5371(80)90379-5)
- Durantini, M. R., Albarracín, D., Mitchell, A. L., Earl, A. N., & Gillette, J. C. (2006). Conceptualizing the influence of social agents of behavior change: A meta-analysis of the effectiveness of HIV-prevention interventionists for different groups. *Psychological Bulletin*, 132, 212–248. <https://doi.org/10.1037/0033-2909.132.2.212>
- Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, 33, 770–782. <https://doi.org/10.3758/bf03193073>

- Ecker, U. K. H., & Ang, L. C. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology, 40*, 241–260. <https://doi.org/10.1111/pops.12494>
- 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*, 185–192. <https://doi.org/10.1016/j.jarmac.2017.01.014>
- 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*, 283–310. <https://doi.org/10.1080/17470218.2010.497927>
- Ecker, U. K. H., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied, 20*, 323–335. <https://doi.org/10.1037/xap0000028>
- 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*, 101–115. <https://doi.org/10.1016/j.jml.2015.09.002>
- Ecker, U. K. H., Lewandowsky, S., Swire, B., & Chang, D. (2011). Correcting false information in memory: Manipulating the strength of misinformation encoding and its retraction. *Psychonomic Bulletin & Review, 18*, 570–578. <https://doi.org/10.3758/s13423-011-0065-1>
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition, 38*, 1087–1100. <https://doi.org/10.3758/MC.38.8.1087>
- Ecker, U. K. H., O'Reilly, Z., Reid, J. S., & Chang, E. P. (2020). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology, 111*, 36–54. <https://doi.org/10.1111/bjop.12383>
- 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. https://doi.org/10.1016/j.jarmac.2020.08.003*
- Ecker, U. K. H., Sze, B., & Andreotta, M. (2021). Corrections of political misinformation: No evidence for an effect of partisan worldview. *Philosophical Transactions of the Royal Society B*. (in press).
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 39*, 175–191. <https://doi.org/10.3758/BF03193146>
- Fein, S., McCloskey, A. L., & Tomlinson, T. M. (1997). Can the jury disregard that information? The use of suspicion to reduce the prejudicial effects of pretrial publicity and inadmissible testimony. *Personality and Social Psychology Bulletin, 23*, 1215–1226. <https://doi.org/10.1177/01461672972311008>
- Ferrero, M., Hardwicke, T. E., Konstantinidis, E., & Vadillo, M. A. (2020). The effectiveness of refutation texts to correct misconceptions among educators. *Journal of Experimental Psychology: Applied. https://doi.org/10.1037/xap0000258*
- Fischer, F. (2019). Knowledge politics and post-truth in climate denial: On the social construction of alternative facts. *Critical Policy Studies, 13*, 133–152. <https://doi.org/10.1080/19460171.2019.1602067>
- Gauchat, G. (2012). Politicization of science in the public sphere: A study of public trust in the United States, 1974 to 2010. *American Sociological Review, 77*, 167–187. <https://doi.org/10.1177/0003122412438225>
- 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*, 104028. <https://doi.org/10.1016/j.jml.2019.104028>
- Gordon, A., Quadflieg, S., Brooks, J. C. W., Ecker, U. K. H., & Lewandowsky, S. (2019). Keeping track of ‘alternative facts’: The neural correlates of processing misinformation corrections. *NeuroImage, 193*, 46–56. <https://doi.org/10.1016/j.neuroimage.2019.03.014>
- 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*, 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*, 201–209. <https://doi.org/10.1016/j.jarmac.2013.10.001>
- Hamilton, L. C., Hartter, J., & Saito, K. (2015). Trust in scientists on climate change and vaccines. *SAGE Open, 5*. <https://doi.org/10.1177/2158244015602752>
- Hinze, S. R., Slaten, D. G., Horton, W. S., Jenkins, R., & Rapp, D. N. (2014). Pilgrims sailing the Titanic: Plausibility effects on memory for misinformation. *Memory & Cognition, 42*, 305–324. <https://doi.org/10.3758/s13421-013-0359-9>
- Hovland, C. I., & Mandell, W. (1952). An experimental comparison of conclusion-drawing by the communicator and by the audience. *Journal of Abnormal and Social Psychology, 47*, 581–588. <https://doi.org/10.1037/h0059833>
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*, 635–650. <https://doi.org/10.1086/266350>
- Ithisuphalap, J., Rich, P. R., & Zaragoza, M. S. (2020). Does evaluating belief prior to its retraction influence the efficacy of later corrections? *Memory, 28*, 617–631. <https://doi.org/10.1080/09658211.2020.1752731>
- Jern, A., Chang, K. M. K., & Kemp, C. (2014). Belief polarization is not always irrational. *Psychological Review, 121*, 206–224. <https://doi.org/10.1037/a0035941>
- 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*, 1420–1436. <https://doi.org/10.1037/0278-7393.20.6.1420>
- Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology, 23*, 77–83. <https://doi.org/10.1016/j.copsyc.2018.01.003>
- Kendeou, P., Butterfuss, R., Kim, J., & van Boekel, M. (2019). Knowledge revision through the lenses of the three-pronged approach. *Memory & Cognition, 47*, 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*, 374–397. <https://doi.org/10.1080/0163853X.2014.913961>
- Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems, 36*, 931–968. <https://doi.org/10.1080/07421222.2019.1628921>
- Krishna, A. (2018). Poison or prevention? Understanding the linkages between vaccine-negative individuals' knowledge deficiency, motivations, and communication behaviors. *Health Communication, 33*, 1088–1096. <https://doi.org/10.1080/10410236.2017.1331307>
- Kuklinski, J. H., Quirk, P. J., Jerit, J., Schwieder, D., & Rich, R. F. (2000). Misinformation and the currency of democratic citizenship.

- Journal of Politics*, 62, 790–816. <https://doi.org/10.1111/0022-3816.00033>
- 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, 1325–1356. <https://doi.org/10.1111/j.1559-1816.2010.00620.x>
- 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, 353–369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Lewandowsky, S., Ecker, U. K. H., Seifert, C., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13, 106–131. <https://doi.org/10.1177/1529100612451018>
- Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science*, 16, 190–195. <https://doi.org/10.1111/j.0956-7976.2005.00802.x>
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12, 361–366. <https://doi.org/10.1101/lm.94705>
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44, 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- McGinnies, E., & Ward, C. D. (1980). Better liked than right: Trustworthiness and expertise as factors in credibility. *Personality and Social Psychology Bulletin*, 6, 467–472. <https://doi.org/10.1177/014616728063023>
- McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The Handbook of Social Psychology* (2nd ed., Vol. 3, pp. 136–314). Reading, MA: Addison Wesley.
- Messing, S., & Westwood, S. J. (2012). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41, 1042–1063. <https://doi.org/10.1177/0093650212466406>
- Moser, S. C. (2010). Communicating climate change: History, challenges, process and future directions. *Wiley Interdisciplinary Reviews: Climate Change*, 1, 31–53. <https://doi.org/10.1002/wcc.11>
- Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science & Medicine*, 211, 274–281. <https://doi.org/10.1016/j.socscimed.2018.06.032>
- Nyhan, B., Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- O’Rear, A. E., & Radvansky, G. A. (2020). Failure to accept retractions: A contribution to the continued influence effect. *Memory & Cognition*, 48, 127–144. <https://doi.org/10.3758/s13421-019-00967-9>
- Paynter, J., Luskin-Saxby, S., Keen, D., Fordyce, K., Frost, G., Imms, C., Miller, S., Trembath, D., Tucker, M., & Ecker, U. K. H. (2019). Evaluation of a template for countering misinformation—Real-world autism treatment myth debunking. *PLOS ONE*, 14, e0210746. <https://doi.org/10.1371/journal.pone.0210746>
- Pornpitakpan, C. (2004) The persuasiveness of source credibility: A critical review of five decades’ evidence. *Journal of Applied Social Psychology*, 34, 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, 11–26. <https://doi.org/10.3758/s13421-013-0339-0>
- 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, 62–74. <https://doi.org/10.1037/xlm0000155>
- Schul, Y., Mayo, R., & Burnstein, E. (2008). The value of distrust. *Journal of Experimental Social Psychology*, 44, 1293–1302. <https://doi.org/10.1016/j.jesp.2008.05.003>
- Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? *Psychology of Learning and Motivation*, 41, 265–292. [https://doi.org/10.1016/S0079-7421\(02\)80009-3](https://doi.org/10.1016/S0079-7421(02)80009-3)
- Sinclair, A. H., Stanley, M., & Seli, P. (2019). Closed-minded cognition: Right-wing authoritarianism is negatively related to belief updating following prediction error. <https://doi.org/10.31234/osf.io/94a7v>
- Smithson, M. (1999). Conflict aversion: Preference for ambiguity vs. conflict in sources and evidence. *Organizational Behavior and Human Decision Processes*, 79, 3, 179–198. <https://doi.org/10.1006/obhd.1999.2844>
- 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>
- Stadler, M., & Bromme, R. (2014). The content-source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). Cambridge, MA: MIT Press.
- Sternthal, B., Phillips, L., & Dholakia, E. (1978). The persuasive effects of source credibility: A situational analysis. *Public Opinion Quarterly*, 42, 285–314.
- Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The role of familiarity in correcting inaccurate information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 43, 1948–1961. <https://doi.org/10.1037/xlm0000422>
- Swire-Thompson, B., Ecker, U. K. H., Lewandowsky, S., & Berinsky, A. J. (2020). They might be a liar but they’re my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, 41, 21–34. <https://doi.org/10.1111/pops.12586>
- Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, 33, 460–480. <https://doi.org/10.1080/10584609.2015.1102187>
- Underwood, J., & Pezdek, K. (1998). Memory suggestibility as an example of the sleeper effect. *Psychonomic Bulletin & Review*, 5, 449–453. <https://doi.org/10.3758/BF03208820>
- 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>
- van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Communication*, 35, 560–575. <https://doi.org/10.1080/10410236.2019.1573295>
- Vasek, L., & Franklin, M. (2010). Penny Wong says beaches could erode away. <https://www.news.com.au/national/wong-says-beaches-could-erode-away/news-story/5f879bb27814e357a8862ee2fa2aa947>
- Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, 39, 621–645. <https://doi.org/10.1177/1075547017731776>
- Wagner, M. F., & Skowronski, J. J. (2017). Social influence and mental routes to the production of authentic false memories and inauthentic false memories. *Consciousness and Cognition*, 51, 34–52. <https://doi.org/10.1016/j.concog.2017.02.017>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85, 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, 155-177. 009365021985460. <https://doi.org/10.1177/0093650219854600>
- Wiener, J. L., & Mowen, J. C. (1986). Source credibility: On the independent effects of trust and expertise. *Advances in Consumer Research*, 13, 306-310.
- Wilkes, A. L., & Leatherbarrow, M. (1988) Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology A*, 40, 361-387, <https://doi.org/10.1080/02724988843000168>
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41, 135–163. <https://doi.org/10.1007/s11109-018-9443-y>
- Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.