Source reliability and the continued influence effect of misinformation: A Bayesian network approach

Jens Koed Madsen (jens.madsen@ouce.ox.ac.uk)
School of Geography and the Environment, University of Oxford
OX1 3QY, South Parks Road, Oxford, United Kingdom
Orcid: 0000-0003-2405-8496

Saoirse Connor Desai (saoirse.connor-desai@city.ac.uk)
Department of Psychology, City, University of London
Rhind Building, St John Street, Clerkenwell, London, UK

Toby Pilditch (t.pilditch@ucl.ac.uk)
School of Geography and the Environment, University of Oxford
OX1 3QY, South Parks Road, Oxford, United Kingdom
&
Department of Experimental Psychology, University College London
26 Bedford Way, WC1H 0AP, London, United Kingdom

Abstract

Misinformation, and its impact on society, has become an increasingly topical field of study of late. A body of literature exists that suggests misinformation can retain an influence over beliefs despite subsequent retraction, known as the Continued Influence Effect (CIE). Researchers have argued this to be irrational. However, we show using a Bayesian formalism why this argument is overly assumptive, pointing to (previously overlooked) considerations of reliability of, and dependence between, misinforming and retracting sources. We demonstrate that lay reasoners intuitively endorse assumptions that demarcate CIE as a rational process, based on the fact misinformation *precedes* its retraction. Moreover, despite using established CIE materials, we further upturn the applecart by finding participants show CIE, and appropriately penalize the reliabilities of contradicting sources.

Keywords: Continued Influence Effect; Negation; Reliability; Dependency; Reasoning

Introduction

Misinformation can have a lasting effect on beliefs that people entertain and on the inferences they can make about events[1]. Poor information, whether spread deliberatively or mistakenly, can have serious and widespread repercussions for society. For example, despite being corrected repeatedly, some people belief that there is a causal link between the measles mumps and rubella (MMR) vaccination and autism.

This belief persists in some communities despite scientific evidence refuting the myth (Horne et al., 2015). Decreased acceptance of the MMR vaccination has contributed to a 7% drop in vaccination rates in the UK and a 1.7-fold increase in refusal to vaccinate in the US (Smith et al., 2008), and consequently, an increase in a vaccine-preventable disease.

The harmful effects of misinformation and ineffectiveness of attempts to correct mistaken beliefs have become a great concern for contemporary society (Gordon, Brooks, Quadflieg, Ecker, & Lewandowsky, 2017; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012), and has recently become a weighty issue for governments, media organizations, and citizens (see Lewandowsky et al., 2017). Problematically, though, studies show that belief in erroneous information can persist even after it has been unambiguously corrected (Lewandowsky et al., 2012)
Regardless of how information is corrected, research shows that it often fails to abolish the effects of misinformation (see Lewandowsky et al., 2012 for review). The so-called *Continued Influence Effect* (CIE) of misinformation refers to the consistent finding that information initially presented as true continues to influence beliefs and reasoning despite clear and credible corrections (Ecker et al., 2011a, 2011b; Johnson & Seifert, 1994; Rich & Zaragoza, 2016).

In the paper, we explore two aspects of CIE. First, no normative account of how people should "optimally" process corrections to misinformation has been provided to date. CIE studies typically report the observed phenomenon in a variety of contexts and settings. To explore the effect systematically, we provide a Bayesian Network model to test whether CIE is truly irrational or if the phenomenon can be explained rationally. Second, past research shows the importance of dependency (Madsen et al., 2018). That is, whether a source is truly independent from another source, or if they are somehow related. This influences the impact of the report on the hypothesis *and* perceived reliability. In accordance with these studies, we manipulate the source of debunking such that the initial source debunks its own statement or a different source debunks the statement.

[1] We define information as any piece of information or evidence that is initially thought to be true, but which later turns out to be erroneous, but which can be corrected. Going beyond the current study, the intention behind the dissemination of misinformation is crucial (e.g. the difference between an honest mistake and a malevolent lie – both of which may provide poor information).

Exploring CIE through a formal reasoning model yields interesting results. First, we find a rational explanation for CIE. We show that belief in the hypothesis remains above prior level, but instead the reliability (in the second reporter case) is penalized. Second, perceived dependence influences the effect. Given a Bayesian network, CIE is irrational only insofar that the sources are entirely *independent* of each other. Comparatively, when considering reports temporally and dependent, CIE is entirely *rational*. Correcting is often done by a source that is, in some way, linked with the initial source of misinformation (e.g. a reporter working at the same network). This highlights a significant conceptual limitation to the way in which CIE is framed classically.

Finally, we can demonstrate irrationality in a manner that is backwards to what is typically reported in CIE studies. In CIE studies, people should not stick with original beliefs given correction, but do so any way. We show cases where there are reasonable grounds for why people should stick with their original beliefs, but do not.

The continued influence effect

Continued influence studies examine corrections to misinformation using variants of a laboratory paradigm first developed by Wilkes and Leatherbarrow (1988; but also see Johnson & Seifert, 1994). There are two leading cognitive explanations for CIE (Gordon et al., 2018; Lewandowsky et al., 2012):

First, ***the selective retrieval account*** argues that CIE occurs when correct and incorrect information are stored in memory simultaneously, and misinformation is activated but inadequately blocked (Ecker et al., 2011a). Second, ***the model updating account*** argues that people continually construct a mental event model as new information becomes available. Correcting information without providing a credible alternative (e.g. a competing causal explanation) leaves people with a gap in their mental model. On this view, people prefer a coherent but incorrect model to a correct but incomplete one and thus maintain the invalidated information (Ecker et al. 2010; Johnson & Seifert, 1994).

A typical CIE task involves a series of sequentially presented statements describing an unfolding event, similar to a breaking news report. Misinformation that allows inferences to be drawn about the outcome of the event is presented early in the sequence, but retracted later. Participants' event comprehension is assessed, typically to show that misinformation continues to influence people's inferential reasoning even though they clearly understand and remember that the information was corrected (Johnson & Seifert, 1994). The effect persists even when given prior warnings about the persistence of misinformation (Ecker et al., 2010). The fact that retractions are often ineffective at 'removing' misinformation from people's understanding of events emphasizes the need to identify and model factors that contribute to the *Continued Influence Effect*.

Sustained reliance on misinformation given a retraction is often depicted as a bias – or systematic deviation from a normative standard – and therefore irrational (e.g. Lewandowsky et al., 2012). This perspective assumes two things; first, that the optimal solution is always to disregard initially prior information in favour of new information, and second that the 'true' value of the retraction is known.

Source reliability

Establishing a source's reliability is critical when deciding whether to rely on the information conveyed to us by other people, and may drive the CIE. Reliability can be separated into issues of: i) observational sensitivity, ii) objectivity, and iii) veracity (Schum, 1994). For example, jurors must establish whether a witness' testimony is truthful and accurate in order to reach a verdict, and voters must similarly place their confidence in the statements of politicians when deciding who to vote for.

While appeals to authority and reliance on testimonies traditionally have been seen as fallacoius (ad verecundiam) or as a shallow cue, Bayesian models have integrated reliability within a normative theory of reasoning (Bovens & Hartmann, 2003; Hahn et al,. 2009; Harris et al., 2015).

People use a range of cues to evaluate a source's reliability. For example, in the legal domain witnesses may contradict themselves or be contradicted by others, which may reassess the credibility (see Connor Desai et al., 2016). Moderating perceived source reliability is an sensible act if new information, additional contradictory or corroborative reports, or insight into whether or not the sources are related to each other is made known. In addition to new infomraiton, source dependecy moderates perceived reliability (Bovens & Hartmann, 2003; Madsen et al., 2018).

Contradiction is particularly relevant to CIE studies where the misinformation and its retraction are typically issued by the same source. A source who announces that they previously gave incorrect information may appear less reliable than one who does not. Consistent with this, one CIE study found that distrust in the source of the retraction was a primary reason for disbelieving the retraction (Guillory & Geraci, 2010; 2013). Indeed, Lewandowsky et al., (2012) argue that source reliability (high and low) may facilitate 'tagging' of correct and incorrect information and facilitate retrieval of information when this information is made salient.

Thus, perceived reliability moderates the degree to which people are willing to integrate reports from more or less reliable sources. If a highly reliable source provides report about an issue, the recepient should *normatively* revise her belief in the suggested direction. Second, reports from independent sources are more diagnostic than reports that stem from sources who share a common background. In order to model reliability estimates, belief in the hypothess, and to develop a formal model of CIE, we adopt a Bayesian approach.

A Bayesian approach to source reliability

As mentioned, CIE studies do not provide a normative account of how people should process retractions to misinformation. The lack of formalism is crucial as there may be situations where continued reliance on misinformation is rational given the lack of information and inherent uncertainty of the situation. In such situations, people may use cues like reliability to assess the validity of misinformation and its retraction, and decide how much to incorporate these pieces of information into their beliefs.

Bayes' theorem gives a normative belief revision model. It integrates people's subjective prior degrees of belief with the likelihood ratio to estimate the posterior degree of belief. It has been applied to conditional reasoning (Oaksford & Chater, 2007), argumentation (Hahn & Oaksford, 2006; 2007), and other areas of cognition (Chater et al., 2010).

To explore CIE formally, we use a Bayesian Network (BN) framework (Pearl, 2000). BNs use graph structures to represent the probabilistic relationships between hypotheses and evidence (including reliability), using conditional probabilities to represent the strength of relations, and show what inferences are rationally permitted from a given model given available information. This is an ideal method for examining whether CIE is rational in some circumstances, as it provides the means to test causal models of scenarios – including their models of the reliability of the sources providing information – and compare inferences to a normative standard (Fenton et al, 2013).

Congruency of information with the misinformation and the reliability of sources providing the misinformation or the retraction are potential moderators of the CIE. BNs provide a formal model to test responses against model predictions and test foundational assumptions of the CIE.

Comparing judgments to Bayesian predictions test if there are situations in which retaining belief in misinformation after a retraction is rational. Formally modelling the causal relations between information included in a scenario would make it possible to test participants' causal models of scenarios. This provides an understanding of the cognitive mechanisms involved in the CIE.

Method

Participants: 101 participants were recruited from Prolific Academic (71 females, age = 31.57±9.6). Participants were paid £1.50 (~$1.97) and took 14 minutes (on average) to complete the experiment.

Stimuli, Design & Procedure: To replicate CIE studies, we used stimuli adapted from past research (Johnson & Seifert, 1994; Gordon et al., 2017, see Table 1 for an example of stimulus material).

Table 1: Example of news report and comprehension probes

Sentence	Control	Retraction (Same Source)	Retraction (Different Source)
Example News Report			
Sentence 1	*A motorcyclist died yesterday after being knocked off his bike by a car.*		
Sentence 2	***Officer Jones reported that the driver of the car had been travelling over the speed limit.***	***Officer Jones reported that the driver of the car was intoxicated.***	***Officer Jones reported that the driver of the car was intoxicated.***
Sentence 3	The accident happened on the A7 north of Carlisle.		
Sentence 4	The motorcyclist was 30 years old and had two children.		
Sentence 5	***Officer Jones revealed that the car driver was not intoxicated.***	***Officer Jones revealed that the car driver was not intoxicated.***	***Officer Smith revealed that the car driver was not intoxicated***
Sentence 6	The driver of the car was also injured in the incident.		
Example Comprehension Probes			
Question 1	Drink-driving charges should be brought against the driver of the car		
Question 2	The driver should be forced to complete a drink-driving awareness course		
Question 3	A breathalyser would have returned a positive result		

It was a between-subjects study with the effect of retracting information was assessed between groups (Control, Retraction – Same Source, Retraction – Different Source). Participants were randomly assigned to a condition.

Sentence 2 differed between control and retraction conditions for each event. In retraction conditions, sentence 2 contained (mis)information. In the control condition, it contained circumstantial information to provide a baseline for the comprehension test. The key sentence (sentence 5) was identical in all conditions. Given exposure to sentence 2, sentence 5 did or did not correct previous information. For source conditions, the source of the (mis)information (sentence 2) and retraction (sentence 5) were either the same (same source) or different (different source).

In all, we tested four scenarios. Presentation order of the scenarios was randomized across participants. The scenarios used were selected from a set of eight pilot reports (N = 70)

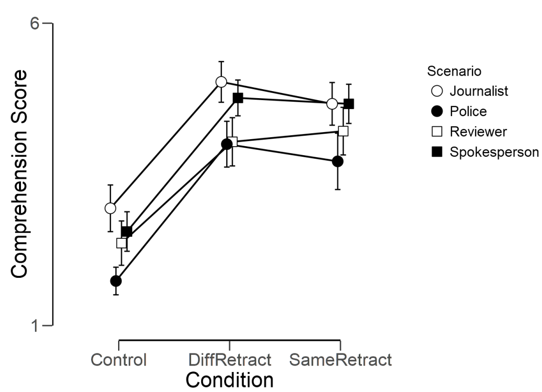

Figure 1. Mean comprehension scores, split by scenario (line) and condition (horizontal axis). Error bars reflect 95% CI.

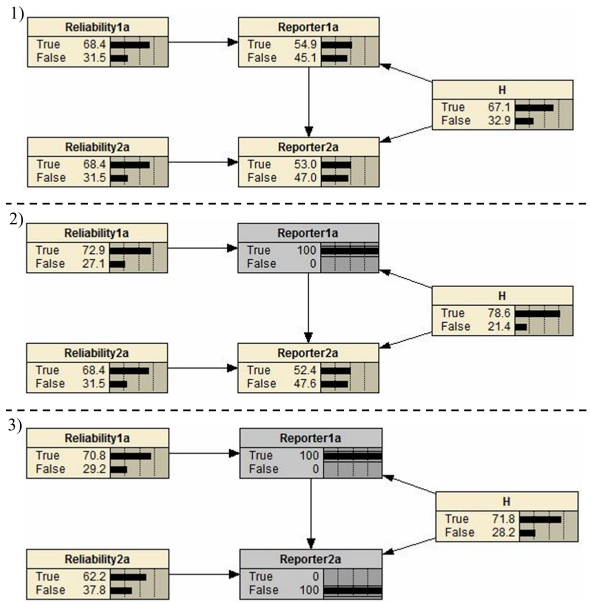

Figure 2a. Group BN model for the retraction different condition, police officer scenario. 1) Baseline (no observation) stage, 2) Single positive (first) report stage (i.e. control condition), and 3) Final (retraction) state given a second, separate reporter.

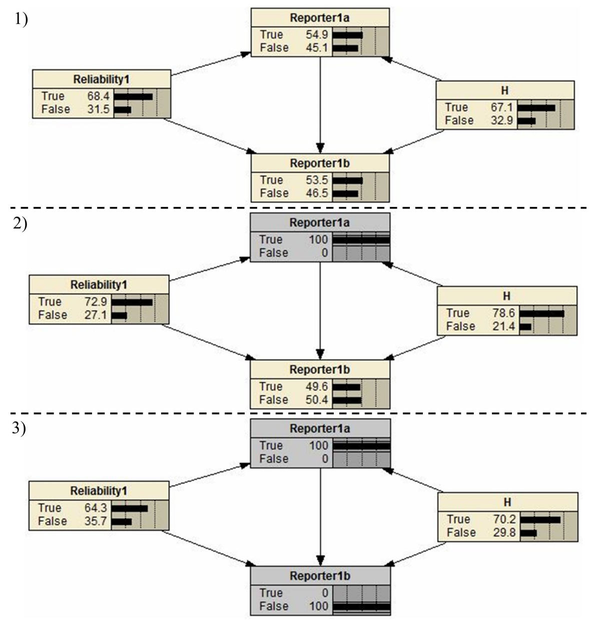

Figure 2b. Group BN model for the retraction same condition, police officer scenario. 1) Baseline (no observation) stage, 2) Single positive (first) report stage (i.e. control condition), and 3) Final (retraction) state given a second, report from the same reporter.

where scenarios with the largest 'continued influence effect' of misinformation were chosen for the actual study.

Prior to reading any scenario, participants provided prior estimates for their beliefs in the reliability of the sources of misinformation that would appear in the subsequent reports and whether they would provide reliable reports. This was measured on a scale of 0 (Extremely unlikely) to 100 (Extremely likely).

Further, to parameterise the model, participants provided six conditional probability estimates per report (24 in total). They rated their belief that the source of report 1 would make an erroneous statement in reporting about an event, if they were or were not reliable on the same scale as used for prior beliefs. Questions about the second reporter differed between the same and different source conditions. Eliciting conditional probabilities allowed for parameter-free models.

Continued reliance of misinformation was measured by a set of comprehension probes that followed each scenario (see Table 1). Participants rated each probe on a 7-point scale from 'strongly disagree' to 'strongly agree'. In line with previous CIE methods, probes referred to the critical information (sentence 5). Higher endorsement of comprehension probes measured the degree to which the misinformation presented in sentence 2 had been incorporated into a participants' understanding of the report.

After rating the probes participants provided their belief posterior probability on a similar scale used for prior beliefs. For example, in the scenario in Table 1, participants were asked: 1) Given everything you know so far about the incident in question, how likely do you think it is that the accident occurred because the driver was intoxicated/travelling over the speed limit? 2) Given everything you know so far about the incident in question, how likely to do you think it is that the police officer is reliable in their reporting? Participants who received a retraction from a different source as the misinformation provided an additional estimate for the reliability of the second reporter.

Results

Bayesian analyses were done with JASP statistical software (JASP Team, 2018) and assumed an uninformed prior.

Comprehension Scores

A Bayesian repeated measures ANOVA was used to determine the effect of condition and scenario type on mean comprehension scores. Strong evidence was found for the main effect of condition, $BF_{Inclusion} = 1.917 * 10^{12}$, and scenario, $BF_{Inclusion} = 5.44 * 10^{9}$, but no interaction, $BF_{Inclusion} = 0.122$. The model including just main effects was the strongest fit, $BF_{M} = 131.26$, and significant overall, $BF_{10} = 2.105 * 10^{22}$. As illustrated in Fig. 1 below, scenarios differed in comprehension scores from one another, and there was a differential influence of condition.

Critically, the effect of condition indicated significantly higher endorsement of comprehension probes following the presentation and retraction of misinformation compared to when no misinformation was presented at all. This indicates that, a CIE was observed across all scenarios, such that a retraction was insufficient to bring endorsement ratings back to baseline.

Bayesian Model Fits

Using the conditional probabilities and priors elicited from participants, group means on these estimates were used to parameterize 2 group-condition models for each scenario. Although the conditional probabilities and priors for each first reporter and reliability node were fitted based on all participants, two important exceptions are noted. First, conditional probabilities for the second reporter were based solely on estimates from the condition of relevance (i.e. only estimates from the retraction different condition were used to parameterize the entailed different second reporter in that condition). Secondly, prior probabilities for each hypothesis were reverse-engineered (via Bayes Theorem) using the posteriors provided by control condition. More precisely, taking the control condition BN model, the posterior for the hypothesis was fitted, given the single positive report. Retracting the observation could reveal the approximate prior (absent observations) for that hypothesis. This "prior" was fitted into the models for the two retraction conditions.

Figs 2a and 2b show example condition models for the Police officer scenario, fitted from participant data according to the protocol outlined above. Several important trends are noticeable:

Firstly, as expected, given a single positive reporter (stage 2), belief in the hypothesis (H) increases, and the predicted likelihood of corroboration from the second report increases. However, when the second, contradicting report is observed (stage 3), the belief in the hypothesis (H) does *not* return to prior (stage 1) levels. Instead, the reliability of sources decreases given the contradiction, this decrease is strongest in the second reporter (different condition), but is also substantial when the same reporter contradicts themselves (Fig. 2b, stage 2 to stage 3).

Critically, the reason for this effect (retention of belief in H, but reduction in perceived reliability) is due to the capturing of the temporal dependence from first to second report. Put another way, the models capture the intuition that a second report is aware of the first report (whether internally in the case of the same reporter condition, or via general narrative in the different reporter condition). The manner and strength of this influence is then captured by the elicited conditional probabilities from participants.

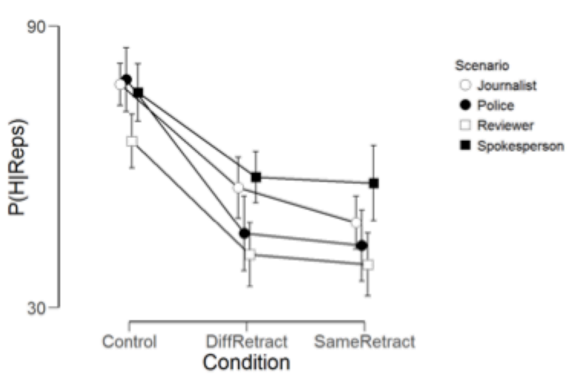

Figure 4. Posterior estimates of belief in the hypothesis (H), given all reports, split by scenario (line) and condition (horizontal axis). Error bars reflect 95% CI

Participant Estimates

Returning to participant data, we again use Bayesian repeated measures ANOVA to examine whether probability estimates correspond to the BN model predictions (and thus map onto a continued influence effect), or corroborate the comprehension score measures (and indicate an absence of CIE – against fitted normative prescription).

Hypothesis. Turning first to posterior estimates of belief in the hypothesis, we find main effects of condition, $BF_{Inclusion} = 3.328 * 10^9$, and scenario, $BF_{Inclusion} = 41812.52$, but no interaction, $BF_{Inclusion} = 0.467$. The model consisting of the main effects along was the strongest fit, $BF_M = 34.27$, and significant overall, $BF_{10} = 2.247 * 10^{14}$. As Fig. 4 illustrates, these effects corroborate comprehension scores, wherein the effect of condition is driven by a reduction in belief in the hypothesis from control to retraction conditions. Crucially, this shows that participants generally deviate from the *prescribed* CIE effect entailed by the BN models, decreasing belief in the hypothesis below the control condition (and prior), given the retraction.

Reliability. Turning next to estimates of reliability, we add to the repeated measures ANOVA analysis a within-subject factor of prior to posterior. Here we find significant main effects of condition (control > retraction different and same), $BF_{Inclusion} > 1.00 * 10^{20}$, scenario, $BF_{Inclusion} = 124.44$, and prior-posterior (posterior < prior), $BF_{Inclusion} > 1.00 * 10^{20}$. Figs 5a-5c illustrate the significant interaction of condition and prior-posterior, $BF_{Inclusion} > 1.00 * 10^{20}$, wherein reliability estimates increased in the control condition (Fig. 5a; where no contradiction occurs, and in line with the increase observed in Fig. 3a and 3b, stage 2), but decreased in both retraction conditions (Fig. 5b and 5c; also in line with model predictions illustrated in Fig. 3a and 3b, stage 3). Lastly, a significant interaction of scenario and prior-posterior was also observed, $BF_{Inclusion} = 75.92$, wherein the spokesperson scenario entailed smaller changes from prior to posterior than the 3 remaining scenarios. The model including the above significant terms yielded the strongest fit, $BF_M = 484.97$, and was significant overall, $BF_{10} = 1.559 * 10^{28}$.

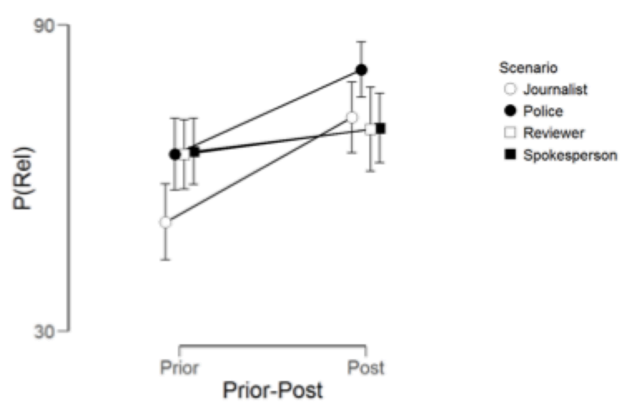

Figure 5a. Control condition reliability estimates for reporters from prior to posterior (reports observed), split by scenario (lines). Error bars reflect 95% CI.

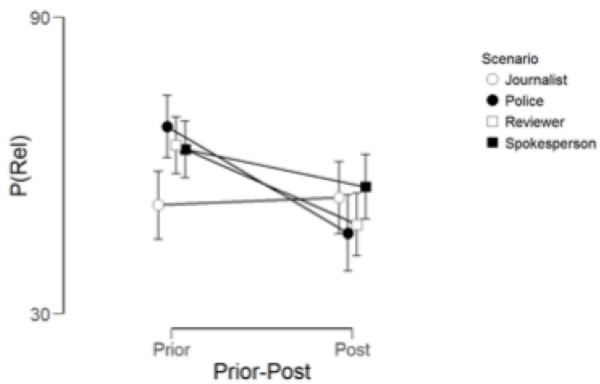

Figure 5b. Retraction different condition reliability estimates for reporters from prior to posterior (reports observed), split by scenario (lines). Error bars reflect 95% CI.

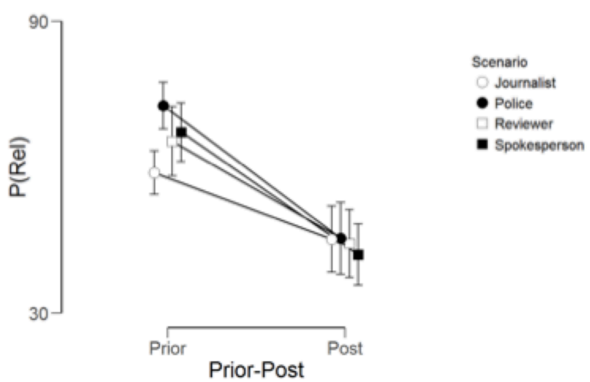

Figure 5c. Retraction same condition reliability estimates for reporters from prior to posterior (reports observed), split by scenario (lines). Error bars reflect 95% CI.

Finally, we note that the retraction condition showed no significant difference in posterior reliability estimates between the two different (first and second) reporters, $BF_{10} = 0.135$, contrary to model predictions (wherein the second reporter should be more substantially penalized).

Discussion and concluding remarks

Previous CIE studies have consistently found that misinformation continues to be influence beyond a clear and credible retraction (e.g. Ecker et al., 2010; 2011a; 2011b; Johnson & Seifert, 1994; Rich & Zaragoza, 2016). Continued reliance on misinformation after a retraction has been depicted as a bias and therefore irrational (Lewandowsky et al., 2012). However, there is an argument that people should exhibit CIE if source reliability judgments are incorporated into how beliefs about misinformation are updated following a retraction.

This paper's aim was to formally model CIE, using a Bayesian Network framework, to capture the temporal dependency between misinformation and its retraction, and the impact this may have on source reliability. We compared participants' judgments to Bayesian predictions to establish whether retaining belief in misinformation (hypothesis) after a retraction is, in fact, sometimes rational.

Participants rated their belief in the hypothesis, and the reliability of sources, when there was no retraction of misinformation, when the retraction was offered by the same source as the misinformation, or by a different source than the misinformation, for a series of news reports.

Behavioural measures showed the standard CIE across all scenarios. Comprehension of the news reports was measured to establish whether misinformation had been incorporated into participant's understanding of the report despite having been retracted. A classic CIE was observed whereby misinformation continued to influence news report comprehension despite being retracted. The effect was observed whether the retraction was offered by the same or a different source to the misinformation.

We also find a rational explanation for CIE. Qualitatively we show that belief in the hypothesis remains above prior level, but instead the reliability of the second reporter (i.e. the retraction) is penalised. Participant's posterior estimates also decreased below their priors, and against what their model predicts. This finding is contrary to the typical account of CIE that people continue to rely on retracted misinformation even though they should. Instead, suggesting that people should continue to rely on misinformation but do not!

Focusing on the condition in which misinformation and retraction come from the same source, participants decrease their estimate for the reporter after they have contradicted themselves, in line with model predictions. In the different source condition, participants decrease their estimates in the reliability of the first reporter (which is incorrect according to the model), and increase reliability estimates of the second reporter (which is correct according to the model). Interestingly, the second reporter was considered more reliable than the first in the police officer and reviewer scenarios (against model predictions), but less reliable than the first in the journalist and spokesperson scenarios (in line with model predictions).

Taking together, we show that participants *should in fact* exhibit a CIE effect (according to fitted Bayesian Network models), and although we find this effect in with standard behavioural measures, we do not observe this with novel probability estimate (P(H) measures. Yet, we do find appropriate penalization in reliability estimates given a contradiction among reports – something hitherto unnoticed in CIE studies, but predicted by our formalism.

To conclude, this research provides a formal account of CIE using the BN framework, and shows that continued reliance on misinformation is in some circumstances rational. This approach captures the qualitative inferences participants make about the reliability of sources of who provide contradictory information. These findings also suggest that perceived reliability moderates the degree to which people are willing to integrate reports from more or less reliable sources.

References

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

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510-516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 811-823.

Cook, J., Ecker, U., & Lewandowsky, S. (2015). Misinformation and How to Correct It. In *Emerging Trends in the Social and Behavioral Sciences* http://doi.org/10.1002/9781118900772.etrds0222.

Connor Desai, S., Reimers, S & Lagnado, D. (2016). Consistency and credibility in legal reasoning: A Bayesian network approach. In Proceedings of the 38th Annual Conference of the Cognitive Science Society, pp. 626-631.

Ecker, U. K., Hogan, J. L. & Lewandowsky, S (2017) Reminders and repetition of misinformation: Helping or hindering its rectraction? *Journal of Applied Research in Memory & Cognition* 6 (2), 185-192

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*(8), 1087–1100.

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

Ecker, U. K., 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*(3), 570–578.

Fenton, N., Neil, M. & Lagnado, D. A. (in press) A General Structure for Legal Arguments About Evidence Using Bayesian Networks, *Cognitive Science*

Gordon, A., Brooks, J. C., Quadflieg, S., Ecker, U. K., & Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing. *Neuropsychologia*, *106*, 216–224.

Hahn, U., Harris, A. J. L., & Corner, A. (2009) Argument content and argument source: An exploration, *Informal Logic 29,* 337-367

Hahn, U., & Oaksford, M. (2006). A Bayesian approach to informal argument fallacies. *Synthese*, *152*(2), 207-236.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological review*, *114*(3), 704.

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2015). The Appeal to Expert Opinion: Quantitative support for a Bayesian Network Approach. *Cognitive Science* 40*,* 1496-1533

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences*, *112* (33), 10321–10324.

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.

Lewandowsky, S., Ecker, U. K., & 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.

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.

Madsen, J. K., Hahn, U. & Pilditch, T. (2018) Partial source dependence and reliability revision: the impact of shared backgrounds*, Proceedings of the 40th Annual Conference of the Cognitive Science Society*

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.

Pearl, J. (2000) *Causality: models, reasoning and inference*, Cambridge University Press

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 & Cognition*, *42*(1), 62–74.

Schum, D.A. (1994). *The Evidential Foundations of Probabilistic Reasoning*, Northwestern University Press.

Smith, M. J., Ellenberg, S. S., Bell, L. M., & Rubin, D. M. (2008). Media coverage of the measles-mumps-rubella vaccine and autism controversy and its relationship to MMR immunization rates in the United States. *Pediatrics*, *121*(4), 836–843.

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.