PAPER IN PRESS AT THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Sensitivity to Misinformation Retractions in the Continued Influence Paradigm: Evidence for Stability

Paul McIlhiney, Gilles E. Gignac, Michael Weinborn, and Ullrich K. H. Ecker

School of Psychological Science, University of Western Australia, WA, Australia

Word count: 7,062 (excluding title page, abstract, references, and online supplement)

Author Note

Paul McIlhiney https://orcid.org/0000-0002-4480-2406

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Paul McIlhiney, School of Psychological Science, University of Western Australia, M304, 35 Stirling Highway, Crawley, W.A 6009, Australia. Email: paul.mcilhiney@research.uwa.edu.au

Abstract

Research has consistently shown that misinformation can continue to affect inferential reasoning after a correction. This phenomenon is known as the *continued influence effect* (CIE). Recent studies have demonstrated that CIE susceptibility can be predicted by individual differences in stable cognitive abilities. Based on this, it was reasoned that CIE susceptibility ought to have some degree of stability itself; however, this has never been tested. The current study aimed to investigate the temporal stability of retraction sensitivity, arguably a major determinant of CIE susceptibility. Participants were given parallel forms of a standard CIE task four weeks apart, and the association between testing points was assessed with an intra-class correlation coefficient and confirmatory factor analysis. Results suggested that retraction sensitivity is relatively stable and can be predicted as an individual-differences variable. These results encourage continued individual-differences research on the CIE and have implications for real-world CIE intervention.

Keywords: misinformation, continued influence, susceptibility, CFA, individual differences

Sensitivity to Misinformation Retractions in the Continued Influence Paradigm: Evidence for Stability

The spread of misinformation has become faster in recent decades, due mainly to the more efficient sharing of information through traditional and social media (e.g., Tsfati et al., 2020; Vosoughi et al., 2018; Wang et al., 2019). This has contributed to the persistence of several misconceptions in modern culture. For example, many still believe that the measles-mumps-rubella (MMR) vaccine causes autism, despite several evidence-based refutations (Taylor et al., 2014); this has contributed to lower rates of MMR vaccination, and thus a greater risk of deadly measles epidemics (e.g., Hoetz, 2016). In the psychological literature, the tendency for misinformation—defined here as any information that is objectively false but presented as true—to affect reasoning and decision making after retraction or correction, is known as the *continued influence effect* (CIE; Johnson & Seifert, 1994). The CIE has received substantial research attention in the last two decades; however, not much is known about how susceptibility to the effect differs between individuals. Moreover, it is unknown whether an individual's susceptibility to the CIE remains stable over time.

In the lab, the CIE is typically elicited using a now standard experimental paradigm (see Chan et al., 2017; Johnson & Seifert, 1994; Walter & Murphy, 2018; Walter & Tukachinsky, 2020; Wilkes & Leatherbarrow, 1988). This paradigm involves giving participants a news report about an incident, which contains a critical piece of information (e.g., what caused the incident). Participants then receive either a retraction of the critical information (retraction condition) or they do not (no-retraction condition). Subsequently, inference questions about the event are given in order to gauge participants' reliance on the critical (mis-)information in their inferential reasoning. For example, Johnson and Seifert (1994) gave participants a news report about a fictional warehouse fire, with the critical information being that the fire had been caused by inappropriate storage of volatile chemicals. Based on answers to the inference questions, Johnson and Seifert determined that participants continued to refer to the volatile chemicals even when a retraction was given—this was despite the participants in the retraction condition stating that they remembered the retraction. Results of this nature are found reliably in CIE research (e.g., Ecker et al., 2017; Walter & Tukachinsky, 2020; also see Rapp & Kendeou, 2007).

The majority of previous lab research on the CIE has sought to understand the effect from two main perspectives, focusing on: (1) the influence of underlying memory processes (e.g., Ecker, Lewandowsky, Chang et al., 2014; Swire, Ecker et al., 2017), and (2) the influence of cognitive biases (e.g., Ecker, Lewandowsky, Fenton et al., 2014; Swire, Berinsky et al., 2017). Using both perspectives, researchers have been able to uncover several aspects related to the mechanisms underlying the CIE. For example, it has been shown that memory-related processes at both encoding (e.g., Ecker, Lewandowsky, Swire et al., 2011) and retrieval (e.g., Swire, Ecker et al., 2017) can contribute to the CIE. Furthermore, cognitive biases have been shown to influence the CIE through perceived source credibility (e.g., Ecker & Antonio, 2020; Guillory & Geraci, 2013), and political beliefs (e.g., Ecker & Ang, 2019; Lewandowsky et al., 2005; Nyhan & Reifler, 2010; but see Ecker et al., 2021; Swire-Thompson, DeGutis et al., 2020; Swire-Thompson, Ecker et al., 2020; Wood & Porter, 2019). However, experiments in both lines of investigation have predominantly manipulated aspects of information presentation in order to observe their impact on the CIE. For example, studies have manipulated study-test delays and the number of misinformation/retraction repetitions (Ecker et al., 2011; Swire et al., 2017), the explicitness of the misinformation (Rich & Zaragoza, 2016), its emotional valence (Chang et al., 2019; Ecker, Lewandowsky & Apai, 2011; Trevors & Kendeou, 2020), whether participants were warned about the CIE prior to misinformation presentation (Clayton et al., 2019; Ecker et al., 2010), or the trustworthiness of the source (e.g., Ecker & Antonio, 2020; Guillory & Geraci, 2013; Swire-Thompson, Ecker et al., 2020). Consequently, the majority of past research has investigated how CIE occurrence changes due to factors associated with the materials, or the congruence of materials with people's worldview.

By contrast, a more recent and less developed line of investigation into the CIE has focused on the influence of individual differences in cognition (Brydges et al., 2018; De keersmaecker & Roets, 2017: Sanderson et al., 2021). The focus of these studies has been to determine whether cognitive differences between people predict CIE susceptibility. For example, Brydges et al. investigated how differences in the capacity of working memory (WM) predicted susceptibility to the CIE, and showed that higher working memory capacity predicted lower CIE susceptibility. It was theorised that these results were due to the fact that successful processing of corrections requires information integration and updating of an event model in memory (e.g., Ecker et al., 2017; Kendeou et al., 2014), which are both functions that rely on working memory (e.g., Singh et al., 2018). De keersmaecker and Roets demonstrated that individual differences in verbal ability—as measured by a modified vocabulary subtest from the *Wechsler Adult Intelligence Scale*—predicted susceptibility to the CIE. Thus, there is evidence that stable traits, such as WM capacity and verbal ability, can help explain people's tendency to show continued reliance on corrected misinformation. However, it should be noted that Sanderson et al. (2021) were unable to replicate the results of Brydges et al. (2019) and showed that the quality of episodic-memory encoding may be a more important determinant of the CIE than WM capacity.

If continued reliance on misinformation is partly predicated on individual characteristics that remain stable over time (e.g., WM capacity; episodic-memory ability), then an individual's CIE susceptibility should also remain relatively stable over time. In other words, someone who relies on corrected misinformation at one point in time should be more likely to rely on (a different piece of) corrected misinformation at a future point in time. One way to test the temporal stability of the CIE would be to evaluate the correlation between two parallel CIE measures administered to an individual at two different time points. If the correlation is relatively large, then it would suggest that CIE susceptibility remains somewhat stable over time; conversely, if the correlation is low in magnitude, then it would suggest that CIE susceptibility is not stable and fluctuates substantially over time.

The Current Study

The current study aimed to investigate whether individual CIE susceptibility remains relatively stable over time or is not predictable as an individual-differences measure. In order to investigate this question, participants were given a standard CIE paradigm task at time 1 and a parallel version of the task four weeks later (time 2). More specifically, the stability of individual differences in retraction sensitivity—arguably, a major determinant of CIE susceptibility—was estimated via intra-class correlation coefficient (ICC) and confirmatory factor analysis (CFA). Specifically, following the correlational analysis, we tested single-factor and correlated two-factor models, with time-1 responses used to represent a time-1 CIE latent variable, and time-2 responses (parallel form) used to represent a time-2 CIE latent variable.[1] The single-factor model was tested in order to ensure that tasks at both time points were measuring the same construct. The correlated two-factor model was tested in order to estimate stability of retraction sensitivity using the CIE latent variables at each time-point. Finally, we formally compared both models, as this can provide stronger evidence for good model fit than the estimates for each individual model (Tomarken & Waller, 2003).

The CIE was measured with the standard methodological approach (e.g., Brydges et al., 2018) to allow for direct comparison with the previous literature. This approach involves determining the difference between a retraction condition (i.e., critical information given and

[1] Note that we continue to use the CIE term in our latent variables, acknowledging that it reflects retraction sensitivity rather than CIE susceptibility per se, to remain consistent with previous studies (i.e., Brydges et al., 2018; Sanderson et al., 2021).

then retracted) and a no-retraction control condition (i.e., critical information given and not retracted); smaller differences between the conditions indicate a smaller effect of the retraction, and thus a larger CIE. Therefore, note again that this approach focuses on the effectiveness of the retraction rather than the continued influence of the misinformation per se. Consequently, our dependent variable was, strictly speaking, *retraction sensitivity*, which could be considered a proxy or strong determinant of more general CIE susceptibility; that is, the more sensitive one is to a retraction, the less likely one is to rely on associated misinformation.[2]

Given that no previous lab research has measured CIE stability in the manner outlined above, there is no precedent for what a "relatively large" test-retest correlation would be. Hedge et al. (2018) recently demonstrated that behavioural tasks designed to produce robust effects in experimental research—such as the CIE task—can show poor reliability in the context of individual-differences (i.e., correlational) research. In fact, their results showed that test-retest reliability of several well-established cognitive measures (e.g., Stroop task, flanker task, go/no-go task etc.) demonstrated intra-class correlation coefficients (ICC) in the range of .00 – .82, with an average reliability of *ICC* = .45; these results were replicated and expanded upon with further cognitive measures by Enkavi et al. (2019). Furthermore, test-retest reliability of related cognitive phenomena (e.g., false memory effects) has varied significantly, ranging from acceptable (e.g., Deese-Roediger-McDermott paradigm, r = .51 -

[2] A separate investigation used an alternative task that potentially focuses more directly on the CIE by measuring *misinformation reliance*, viz. the difference between a retraction condition and a no-misinformation baseline. The use of this task was based on Gordon et al. (2019), who suggested it may be another viable way to measure the CIE, despite potential difficulties establishing the baseline. However, this task demonstrated poor internal-consistency reliability, and a Keiser-Meyer-Olkin (KMO) value suggesting that the data were not appropriate for factor analysis (*KMO* = .52 < .65; Kaiser & Rice, 1974). Consequently, analyses in this investigation were conducted only to try help identify the issues with the task, so that future research may be better informed; for this reason, and for the sake of transparency (e.g., see Franco et al., 2014), we report this investigation in the Online Supplement at https://osf.io/nj9kr.

.76; Blair et al., 2002), to poor (e.g., semantic-priming paradigm, r = .00 - .29; Heyman et al., 2018). Considering this, and the fact that we were only investigating whether CIE susceptibility had *any* demonstrable stability, a cut-off criterion of ICC = .45 (based on the average reliability found by Hedges et al., 2018) was used in the current study to signify relative stability.[3]

Based on Brydges et al. (2018) and De keersmaecker and Roets (2017), it was hypothesised that retraction sensitivity, as a proxy of CIE susceptibility, would remain relatively stable over time (H1). Specifically, it was predicted that there would be a significant (observed-score) correlation of $ICC \geq .45$ between CIE measures at time 1 and time 2, with a CFA demonstrating good model fit for both a single-factor model (including time-1 and time-2 measures) and a correlated two-factor model; the correlated two-factor model was expected to be superior in a formal model comparison.

Method

This study involved presenting a standard CIE task at two time points four weeks apart. The primary aim of the study was to evaluate temporal stability of retraction sensitivity and, by extension, CIE susceptibility; this was based on both the correlation between the time-1 and time-2 CIE task scores, as well as the associated CFA model fits. The research was approved by the University of Western Australia's Human Research Ethics Office.

Participants

Participants were U.S.-based Amazon Mechanical Turk (MTurk) workers, who were recruited and tested via CloudResearch (Litman et al., 2017). Since a minimum of 200 participants is recommended for the use of CFAs (Boomsma & Hoogland, 2001), and exclusions were anticipated, we recruited 300 participants. Participants were excluded from

[3] Note that Enkavi et al. (2019) provided a median reliability for their tasks instead of a mean reliability (*median ICC* = .31), and so an average could not be used from their data to further inform our cut-off criterion.

analysis based on five criteria.[4] Specifically, participants were excluded if they (1) did not complete an English competency task to a satisfactory level; (2) incorrectly answered more than four of the six basic memory questions included in the experiment; (3) indicated that they had not put adequate effort into the task (see Materials and Procedure sections for details); (4) had a response SD < 0.5 across all inference-question Likert-scale scores; or (5) had a mean response SD > Q3 + 2.2 * IQR (Hoaglin & Iglewicz, 1987) across all inference-question sets, indicative of erratic responding.

Twenty-four participants were excluded based on the aforementioned criteria, and 26 did not return for testing at time 2. The final sample size for analysis was thus N = 250 (141 females, 108 males, 1 of undisclosed gender; mean age M = 41.28, SD = 12.68; age range: 21-79). The experiment took approx. 15 minutes to complete at each time-point. Participant payments were US$2.40 and US$3.10 at times 1 and 2, respectively.

Materials

The CIE task was implemented using Qualtrics software (Qualtrics, Provo, UT). Twelve event reports and accompanying questionnaires were presented in total, with half given at each time-point. Test questionnaires were given at the end of the task, once all reports had been presented, and followed the presentation order of the reports. Each report described a different event and contained some critical information relating to the event's cause. Each report existed in retraction and no-retraction control versions. The presentation order always started with a control report and then alternated between conditions (i.e., control-retraction-control-retraction etc.); this was done to reduce build-up of participant expectations regarding the provision of retractions. Moreover, the presentation order of the

[4] Criteria (4) and (5) were added post-hoc based on reviewer feedback. Conclusions were not affected by these additional criteria, however, results differed slightly from the original manuscript. Results using the original a-priori exclusion criteria from the original manuscript are available at https://osf.io/nj9kr.

reports and assignment of reports to conditions was counterbalanced using a Latin square (see Table S1 in the Online Supplement, available at https://osf.io/nj9kr/).

Event Reports

Each report consisted of two separate articles that were roughly 100 words each. The first article contained the critical information, while the second article contained either a retraction of the critical information (retraction condition), or further neutral information about the event (no-retraction control condition). For example, one report described a wildfire, and suggested that the cause of the fire was arson, which was subsequently retracted in the retraction condition but not in the control condition. Further information about event reports can be found in Table 1, and all reports are provided in the Online Supplement.

Questionnaires

Test questionnaires comprised one memory and four inference questions per report, following precedent (e.g., Brydges et al., 2018). The multiple-choice memory questions were provided to ensure participants had encoded the reports (e.g., "How many acres of bushland were burnt? – a. 100,000; b. 50,000; c. 25,000; d. 200,000"). The inference questions were designed to measure reliance on the critical information. For each report, one inference question used a multiple-choice format (e.g., "What do you think was the main cause of the fire?" – a. accident; b. extreme heat; c. arson; d. lightning; e. none of the above); three questions asked participants to rate their endorsement of a statement, using an 11-point Likert scale (e.g., "Malicious intent contributed to the fire" – *strongly disagree* [0] to *strongly agree* [10]). Further information about questionnaires can be found in Table 1, and all questionnaires are provided in the Online Supplement at https://osf.io/nj9kr.

Table 1

Topic, Critical Information, and Example Inference Question for all 12 Scenarios Presented to Participants.

Topic	Critical Information (Event Cause)	Example Inference Question[a]
1. Airplane emergency landing	Severe weather	"The accuracy of pre-flight weather reports should be scrutinized"
2. Wildfire	Arson	"Malicious intent contributed to the fire"
3. Death of a drug dealer	Assault	"'Dealer dead after drug deal gone wrong' would be an appropriate headline for this report."
4. Collapse of woman at bar	Drink spiking	"Police should further investigate the circumstances of the woman's collapse."
5. Train derailment	Speeding	"Dangerous actions by the driver contributed to the derailment"
6. Fish deaths	Chemical waste dumping	"Chemical contamination contributed to the incident"
7. Explosion at warehouse	Improper storage of volatile chemicals	"The explosion was as a result of negligence."
8. Burglary	Organised gang crime	"Local residents should be wary of local gang activity"
9. Car crash	Drunk driving	"This crash could be used to support local drink driving campaigns."
10. Data leak	Employee misconduct	"'Employee leaks sensitive user data' would be an appropriate headline for the report."
11. Mass food poisoning	Breach of food safety guidelines	"The café should be fined for negligent practices."
12. Soccer players' suspension	Involvement in match fixing scheme	"The accused players should not be trusted to play for the team in future."

[a] All inference questions were answered on a Likert scale from strongly disagree (0) to strongly agree (10).

CIE Calculation

In order to determine reliance on the misinformation, inference scores were derived from responses to the inference questions. Responses to the Likert-scale inference questions were divided by 10, converting them to a 0-1 scale, while responses to the multiple-choice inference questions were coded as either 0 (no misinformation reliance in retraction condition; i.e., chose a non-misinformation option) or 1 (misinformation reliance in retraction condition; i.e., chose the misinformation option). A participant's average inference score for each report was then calculated using all four relevant items on the new 0-1 scale (i.e., three rating-scale items and one multiple-choice item). Three difference scores were calculated by pairing retraction and no-retraction scores by order (i.e., subtracting the inference score of the first [second, third] control report from the inference score of the first [second, third] retraction report).[5] These three difference scores formed the three observed scores that served as indicators of the CIE latent variables in our CFAs. Finally, a mean was calculated from the three difference scores, producing an observed CIE score. A larger (negative) score indicated a stronger retraction—and therefore a smaller CIE.[6]

Procedure

At time 1, participants were initially provided with an ethics-approved information sheet and gave informed consent before starting the experiment. All participants then completed an English competency task, which involved writing 3-4 sentences about a subject of their choosing within a 60-second time limit.[7] At each time point, participants were given six event reports and were instructed to read them carefully, as they would be answering

[5] Note that the specific, arbitrary way in which the retraction and control conditions were paired had no impact on the results.

[6] Analyses using the raw retraction and control inference scores were also run, and results can be found in the Online Supplement.

[7] Thank you to Prof Andy Perfors of the University of Melbourne for providing us with the English competency task materials, available at https://qualification-test.appspot.com/.

questions related to the reports later. The articles within each report were presented for a minimum of 15 seconds each, after which participants were able to click to the next section. After encoding of the reports, participants completed a one-minute distractor task (a word sleuth), following precedent (e.g., Brydges et al., 2018). Subsequently, participants were given the test questionnaires. Four weeks after completing the first part of the experiment, participants received an e-mail inviting them to complete the second part. Part 2 was open for three days, but participants were encouraged to complete it within 24 hours of e-mail receipt.

Data Analysis

All ICCs and their 95% confidence intervals (CIs) were tested in SPSS 26 using the reliability analysis function, and followed Koo and Li's (2016) recommended parameters for test-retest reliability analyses—that is, a two-way mixed-effects model looking for absolute agreement. CFAs were tested in Amos 26 (Arbuckle, 2019) using maximum-likelihood estimation. Parameter-estimate CIs were estimated with bootstrapping (2,000 samples). Model fit was evaluated based on the following close-fit criteria (Schweizer, 2010): comparative fit index (CFI) ≥ .950; Tucker-Lewis index (TLI) ≥ .950; standardised root mean-square residual (SRMR) < .08; root mean square error of approximation (RMSEA) < .06 (including 90% CIs). We report implied model χ^2 statistics for completeness. The formal CFA model comparison was based on the following criteria: TLI difference > .010 (Gignac, 2007); Bayesian information criterion (BIC) difference > 2.00 (Raftery, 1995; lower BIC values indicate better fit); and Akaike information criterion (AIC) difference > 2.00 (Burnham & Anderson, 2002; lower AIC values indicate better fit).

Results

Internal-Consistency Reliability and Descriptive Statistics

Prior to testing the single-factor and correlated two-factor models of time-1 and time-2 CIE indicators, we estimated internal-consistency reliability (coefficient ω; following

Gignac, 2014) associated with the CIE latent variables at each time point. As can be seen in Table 2, the CIE tasks demonstrated satisfactory internal consistency for same-condition inference scores (i.e., scores within retraction or control conditions, respectively); however, internal consistencies for retraction – control difference scores (which form the CIE latent variable) were somewhat lower. It has been commonly observed that difference scores exhibit lower reliability than their component scores (e.g., Caruso, 2004; Cronbach & Furby, 1970; Hedges et al., 2018). However, Hedges et al. (2018) argued that this should not eliminate difference scores as a viable method of measurement, as it is important in some contexts to subtract a baseline measure in order to control for confounding effects; this is true in the context of a CIE task. Further, some empirical evidence has shown that difference scores can be sufficiently reliable (e.g., Thomas & Zumbo, 2012; Trafimow, 2015). Moreover, the current study utilised CFAs, which can potentially control for reliability issues with observed scores; appropriately, the Keiser-Meyer-Olkin (KMO) value we calculated for time-1 and time-2 difference scores suggested that our data were suitable for factor analysis ($KMO = .78 \geq .65$; Kaiser & Rice, 1974). As shown in Table 2, time-1 and time-2 CIE scores exhibited low levels of skew |<1.0| and kurtosis |<1.0|, suggesting a suitably-normal distribution for ICC analysis (Bishara & Hittner, 2012).

Table 2

Internal-Consistency Reliability and Descriptive Statistics for CIE Scores.

	Same-condition ω[a]	Difference-score ω	Difference-score ω 95% CI	*M*	*SD*	Skew	Kurtosis
t_1	.76	.53	.40 – .61	-.36	.31	.12	-.55
t_2	.78	.60	.50 – .67	-.33	.30	-.03	-.82

Note. t_1, time 1; t_2, time 2; [a] Estimates were calculated as the mean of ω estimates of the retraction and control conditions, so 95% CIs were not available.

Preliminary Analyses

A manipulation check was performed by determining whether there was a statistically significant difference in inference scores between retraction and control conditions. The difference between retraction (M_{t1} = .40, SD_{t1} = .25; M_{t2} = .41, SD_{t2} = .27) and control (M_{t1} = .76, SD_{t1} = .17; M_{t2} = .74, SD_{t2} = .16) conditions was significant at both time 1, $t(249) = 18.25$, $p < .001$, Cohen's $d = 1.15$, and time 2, $t(249) = 17.23$, $p < .001$, $d = 1.09$, in the expected direction. A check for the presence of a CIE was also performed by determining whether there was a statistically significant difference between retraction-condition scores and zero (following precedent; e.g., Ecker et al., 2017). One-sampled *t*-tests revealed a significant difference between all retraction-condition scores and zero at both time 1, $t(120\text{-}128) \geq 10.87$, $p < .001$, and time 2, $t(122\text{-}126) \geq 9.18$, $p < .001$, indicating the presence of a CIE at both time points.

Intra-class Correlation Coefficient (ICC) and Confirmatory Factor Analysis (CFA)

In order to estimate stability of CIE scores, we first estimated an observed-score ICC across time 1 (M = -.36, SD = .31) and time 2 (M = -.33, SD = .30), ICC = .47, $p < .001$ (95% CI = .37 – .56). This firstly signified the presence of stability in retraction sensitivity; further, it satisfied our a-priori criterion for adequate stability ($ICC \geq .45$; Hedges et al., 2018), and indicated that 22% of variance was shared between time-points.

Next, a single-factor model defined by both time-1 and time-2 indicators was constructed and found to be associated with excellent model fit, $\chi^2 = 11.635$, $p = .235$, CFI = .986, TLI = .976, $SRMR$ = .034, $RMSEA$ = .034 (90% CI = .000 – .084), suggesting that time-1 and time-2 tests measured the same construct. As can be seen in Figure 1, all parcels loaded positively. Furthermore, all factor loadings were significant ($p < .001$).

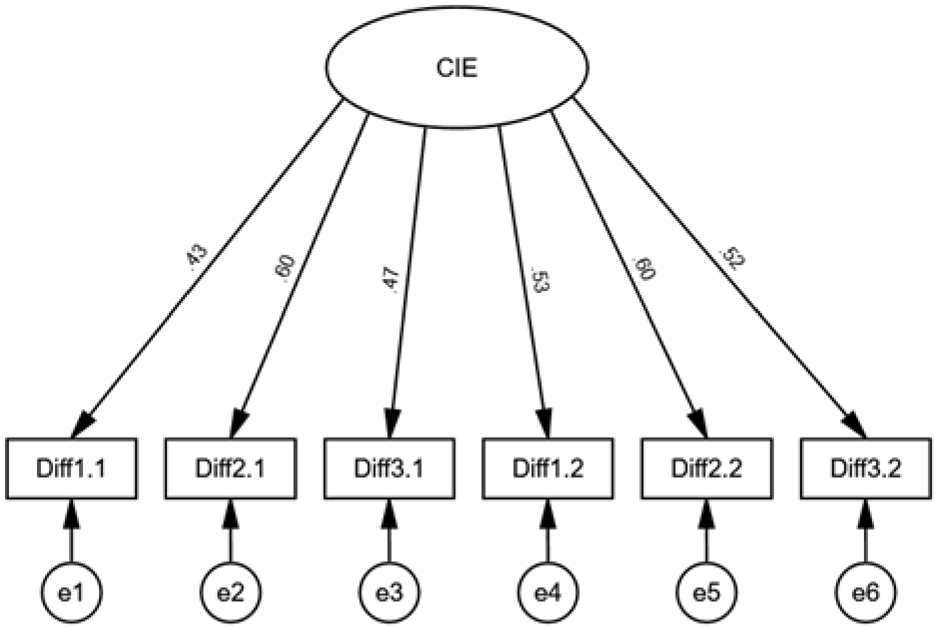

Figure 1. Structural equation model of the CIE latent variable with time-1 (x.1) and time-2 (x.2) indicators. Straight lines with single arrows are regression paths. The observed variables at the bottom (Diffx.x) represent the retraction – control difference scores 1-3. Error terms associated with each observed variable are indicated by e1-6.

Subsequently, a correlated two-factor model with separate CIE latent variables at times 1 and 2 (CIE1, CIE2) was tested, which also showed excellent model fit, $\chi^2 = 7.397$, $p = .495$, $CFI = 1.000$, $TLI = 1.006$, $SRMR = .027$, $RMSEA = .000$ (90% $CI = .000 – .071$), suggesting high stability in the retraction sensitivity construct across time points. Furthermore, as can be seen in Figure 2, there was a large correlation between the time-1 and time-2 CIE latent variables ($r = .83$, $p < .001$, 95% $CI = .64 – 1.00$). In addition, all parcels loaded positively and significantly ($p < .001$) on the respective CIE latent variables.

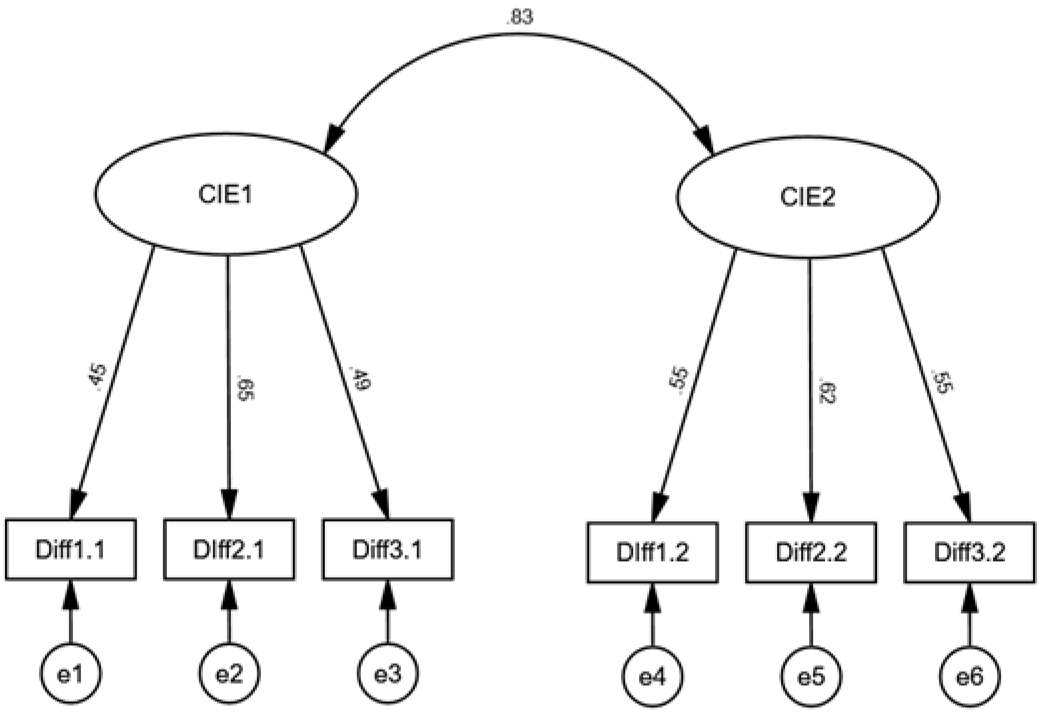

Figure 2. Structural equation model of the CIE latent variables at time 1 (CIE1) and time 2 (CIE2) with time-1 (x.1) and time-2 (x.2) indicators. Straight lines with single arrows are regression paths. The curved line with double arrows is a correlation. The observed variables at the bottom (Diffx.x) represent the retraction – control difference scores 1-3. Error terms associated with each observed variable are indicated by e1-6.

Finally, in order to determine the preferred model for our data, TLI, BIC, and AIC values of the single-factor and correlated two-factor models were compared. The correlated two-factor model demonstrated a better (i.e., higher) TLI value ($TLI_{2\text{-factor}}$ = 1.006; $TLI_{1\text{-factor}}$ = .976, $TLI\Delta$ = .030), and better (i.e., lower) AIC value ($AIC_{2\text{-factor}}$ = 33.397, $AIC_{1\text{-factor}}$ = 35.635, $AIC\Delta$ = 2.238). BIC was uninformative, with the observed difference not satisfying our a-priori criterion ($BIC_{1\text{-factor}}$ = 77.893, $BIC_{2\text{-factor}}$ = 79.176, $BIC\Delta$ = 1.283). Thus, the correlated two-factor model appears to be the preferred model for our data. However, it should be noted that (1) the single-factor model chi-square was not statistically significant (χ^2 = 11.635, p = .235), and (2) the upper-bound 95% CI for the latent-variable correlation in the correlated two-factor model was r = 1.00 (p < .001). Considering this, and the greater simplicity of the single-factor model, one could also argue that the single-factor model should be preferred.

Discussion

The current study aimed to investigate whether an individual's susceptibility to the continued influence effect (CIE) is relatively stable over time or cannot be predicted as an individual-differences measure. Participants were given parallel versions of a conventional CIE task (e.g., Brydges et al., 2018) four weeks apart, and the association between CIE scores at each time point was assessed with a confirmatory factor analysis (CFA). It was hypothesised that retraction sensitivity—a proxy for CIE susceptibility—would remain relatively stable over time, and would thus demonstrate a significant test-retest correlation over a four-week period ($ICC \geq .45$), as well as good model fit in a CFA. This hypothesis was supported by the current study's results.

The significant, positive (observed-score) intra-class correlation coefficient found between time-1 and time-2 CIE scores ($ICC = .47$) provides initial evidence that retraction sensitivity has some temporal stability, and can thus be predicted as an individual-differences variable. Further evidence was provided by the excellent model fit found for single-factor and correlated two-factor models of time-1 and time-2 CIE indicators. The excellent model fit found for the single-factor model indicates that the CIE tasks at each time point were measuring the same construct (i.e., retraction sensitivity), which could be taken as evidence of good validity for the CIE task. Further, the excellent model fit and high latent-variable correlation ($r = .83$) found for the correlated two-factor model suggest that the underlying construct of retraction sensitivity remained highly stable across time points. While our formal model comparison did not determine decisively which model was preferable, we can suggest that both models provide some validation of the retraction sensitivity construct and its partially stable nature.

The present results may also provide evidence for stability in more general susceptibility to the CIE, inasmuch as greater sensitivity to a retraction can be assumed to yield reduced CIE susceptibility. This relationship may manifest itself through stronger encoding of retractions in memory. Specifically, it has been theorised and empirically demonstrated that more strongly encoded retractions reduce the CIE to a greater degree (e.g., Ecker et al., 2017; Ecker et al., 2011). Furthermore, it has been suggested that the salience of a retraction may influence its effectiveness (e.g., Kendeou et al., 2014). Thus, a person who is more sensitive to a retraction may benefit more from a retraction's salience and/or may encode corrective information more strongly, thus reducing subsequent reliance on misinformation. It should be noted, however, that an individual's initial belief in the misinformation (e.g., gullibility) is also critical to overall CIE susceptibility. While we attempted to directly assess the stability of misinformation reliance with an alternative CIE task, these analyses encountered psychometric issues (see the Online Supplement), meaning that strong conclusions could not be drawn. Therefore, future research might aim to address psychometric issues with the alternative CIE task to assess the temporal stability of misinformation reliance relative to a no-misinformation baseline more directly. Nevertheless, based on the above, it seems likely that temporal stability in retraction sensitivity is indicative of temporal stability in CIE susceptibility.

Consequently, the present results corroborate recent CIE research that has investigated the role of individual differences in cognitive abilities, demonstrating that misinformation reliance can be predicted by working memory capacity (Brydges et al., 2018), verbal ability (De keersmaecker & Roets, 2017), or the fidelity of episodic encoding or general episodic-memory ability (Sanderson et al., 2021). If retraction sensitivity had shown complete temporal instability, these findings would have been harder to interpret, and may have implied that findings were context-specific. The current study's results suggest that there is a plausible and consistent role for stable cognitive predictors of the CIE, thus providing further justification for future individual-differences research on the CIE.

The current results also provide some support for current cognitive theoretical accounts of the CIE; specifically the mental-model-updating (e.g., Ecker et al., 2017; Kendeou et al., 2014; Rapp & Kendeou, 2007) and retrieval accounts (e.g., Ayers & Reder, 1998; Ecker et al., 2011; Ecker et al., 2010). Briefly, the mental-model-updating account proposes that the CIE originates from a failure to integrate a correction into memory and update the associated mental model of an event; conversely, the retrieval account suggests that the CIE originates from automatic retrieval of misinformation (e.g., based on its familiarity) and concurrent failure to engage strategic retrieval of a correction. The identification of temporally-stable variance in the CIE provides some validity for both theoretical accounts, as an effect predicated on influences from stable memory processes should itself demonstrate some of said stability.

Concordantly, other stable traits may be of interest with regards to the CIE. For instance, individual differences in executive function—a category of higher cognitive abilities that contains potentially CIE-relevant subcomponents (see Miyake et al., 2000)—could help to explain individual variation in propensity for the CIE. This proposal aligns with the findings of Brydges et al. (2018), as working memory updating is a subcomponent of executive function; moreover, the integration of new information into memory and subsequent updating of an existing mental model (i.e., the role of working memory updating) is central in CIE theory (e.g., Ecker et al., 2017; Kendeou et al., 2014). Thus, investigating the potential links between CIE susceptibility and executive function may help to improve theoretical understanding of the CIE.

Furthermore, CIE research on stable traits could investigate the potential influence of personality traits—which has yet to be investigated. For example, given that scepticism can reduce reliance on misinformation (e.g., Lewandowsky et al., 2005; Rapp et al., 2014) and scepticism can be induced by negative mood (Forgas & East, 2008), it is plausible that trait neuroticism could relate to proclivity for the CIE. More specifically, those who are high in trait neuroticism (i.e., who experience more negative emotion) may be more predisposed to sceptical thinking, and therefore less susceptible to the CIE. Further, it could be speculated from CIE theory that trait conscientiousness could relate to the CIE. The retrieval account of the CIE suggests that a strategic retrieval process is an important antecedent to successful suppression of misinformation use (Ayers & Reder, 1998; Ecker et al., 2011); strategic retrieval is cognitively demanding, and consequently, those high in trait conscientiousness may be more inclined to engage in strategic retrieval and may thus demonstrate lower propensity for the CIE. In summation, then, the current results provide new opportunities to understand the CIE from the perspective of individual differences in stable traits and abilities.

However, it would be misleading to claim from the current results that CIE susceptibility is *largely* stable. Most immediately, the test-retest correlation in the present study indicated that only 22% of the variance in observed CIE scores was shared between both time points. Furthermore, several past studies have demonstrated that the CIE is sensitive to external factors, such as plausibility manipulations (e.g., Baadte & Dutke, 2012; Hinze et al., 2014; Lewandowsky et al. 2005; Lombardi et al., 2016; Rapp & Kendeou, 2007; Seifert, 2002) and the factors mentioned in the Introduction, including the presence of misinformation warnings (Clayton et al., 2019; Ecker et al., 2010), and the repetition of corrections (Ecker et al., 2011). However, it has proven difficult to eliminate the CIE even with strong interventions (Ecker & Antonio, 2020; Ecker et al., 2010; Rapp et al., 2014), which may be partly indicative of the stable portion of CIE susceptibility.

The external factors outlined above can also interact with people's cognitive biases based on their pre-existing beliefs and attitudes, which have also been shown to influence how people use (mis-)information in their reasoning (e.g., Ecker et al., 2014; Swire, Berinsky et al., 2017). Such effects are not entirely avoidable, due to the semantically rich materials that are required in CIE paradigms. Specifically, people's prior knowledge and experience is likely to inform their judgements in a task with semantically rich materials (like the CIE paradigm). For example, a belief that drunk driving is highly prevalent, perhaps from personal experience, could encourage higher endorsement of a claim that drunk driving caused an accident—potentially even after correction. In other words, some topics will resonate more (or less) with people, which could influence the effectiveness of corrections, and therefore add variance. This makes it highly unlikely that the CIE can be entirely stable.

In consequence, it will be important for future research to further investigate the relationship between cognitive (i.e., stable) and contextual (i.e., non-stable) determinants of the CIE, and how this relationship influences temporal stability. Some initial insight could be gleaned from current theoretical accounts of the CIE. Specifically, different theoretical CIE accounts would place different weights on several previously established cognitive and contextual factors. For example, the mental-model-updating account would suggest that individual differences in memory integration and working memory updating abilities matter most, and that ease of conflict detection (i.e., salience of a discrepancy) may be the most influential contextual factor (e.g., Brydges et al., 2018; Ecker et al., 2017; Gordon et al., 2017; Kendeou et al., 2014). Conversely, the retrieval account would suggest that individual differences in episodic long-term memory ability may be important, and that additional variance from context would most likely be introduced by correction distinctiveness or retention-interval duration (e.g., Ecker et al., 2010; Sanderson et al., 2021; Swire, Ecker et al., 2017). However, both CIE accounts would see correction plausibility and source credibility as important contextual factors. While the present study was not designed to further elucidate the specific cognitive determinants of retraction sensitivity and CIE susceptibility, it demonstrates that these determinants are important, and the CIE is unlikely to be fully determined by contextual factors. Additional research is needed to improve our understanding of the origins of stability in the CIE.

However, it could be argued that the observed-score correlation in the current study underestimates CIE stability. Specifically, it was found that: (1) internal-consistency reliability was mediocre for retraction – control difference scores at each time point (i.e., the difference scores that were averaged to produce the CIE scores), and (2) the correlation between time-1 and time-2 CIE latent variables in the CFA was much higher (r = .83) than the observed-score correlation (ICC = .47). These two findings in combination could suggest that the lower internal consistency of the CIE task was limiting the observed-score test-retest correlation. Indeed, it is known that the reliability of two tasks can limit the possible correlation between them (Nunnally, 1970). Thus, a limitation of the present study is the limited reliability of the CIE tasks. While CFAs can be used to control for this limitation—as per the current study—future CIE studies investigating individual differences should endeavour to improve reliability in order to achieve more robust observed-score CIE estimates. One potential way of doing this could be to increase the number of items in the conventional CIE task, as the present version used only three reports per condition and time-point (and thus had only three retraction – control difference scores per time-point). While a longer task may place higher demand on participants' memory, it could also produce more robust observed CIE scores.

Further to the above, the only previous CIE studies to our knowledge to report internal-consistency reliability of the standard CIE task are Brydges et al. (2018) and Sanderson et al. (2021); thus psychometric data on the conventional CIE task is scarce (as was also noted by Swire-Thompson, DeGutis et al., 2020). This is likely due to the experimental nature of most previous CIE studies, which were interested in reliably producing the CIE to investigate the impact of experimentally manipulated contextual factors. By contrast, Brydges et al., Sanderson et al. and the present study are individual differences (i.e., correlational) studies, wherein the internal-consistency reliability of items is more important. In the current study, estimates of internal consistency were acceptable for retraction and control-condition scores ($\omega_{t1} = .76$, $\omega_{t2} = .78$), but were mediocre for retraction – control difference scores ($\omega_{t1} = .53$, $\omega_{t2} = .60$). Similarly, Brydges and colleagues reported internal consistency of $\alpha = .654$ for retraction – control difference scores, while Sanderson and colleagues reported an estimate of only $\alpha = .460$. Although reliability issues with difference scores are well documented (e.g., Caruso, 2004; Cronbach & Furby, 1970; Hedges et al., 2018), CIE tasks need to compare performance in a retraction condition to some baseline in order to draw conclusions about retraction sensitivity or reliance on corrected misinformation directly. Thus, we suggest that future studies—especially individual differences studies—use a latent-variable approach for calculating the CIE, in order to help control for reliability issues with difference scores.

Regarding the application of the present findings, recommendations can be made for real-world CIE intervention strategies—such as those in *The Debunking Handbook 2020* (Lewandowsky et al., 2020). Current intervention strategies largely focus on manipulating contextual factors that influence CIE occurrence (e.g., Ecker et al., 2010; Seifert, 2002). Unfortunately, such intervention strategies can sometimes be hard to implement in the real world (Lewandowsky et al., 2012). Moreover, interventions that target information and its presentation may be relatively ineffective for those who are inherently more susceptible to the CIE or less sensitive to retractions. However, the stable portion of susceptibility to the CIE may not necessarily be resistant to change. Thus, based on the current findings, individual-level interventions designed to target individual differences in susceptibility—such as education programs that teach people about misinformation effects, reasoning strategies, and logical fallacies—could be recommended to help supplement context-focused interventions, and further reduce misinformation impact. In fact, it has already been demonstrated that education-focused interventions can be effective (e.g., Basol et al., 2020; Bonetto et al., 2018; Cook, 2017; Cook et al., 2014; Cook et al., 2017; Maertens et al., 2020; Osborne, 2010; Roozenbeek & van der Linden, 2019; van der Linden et al., 2017).

Limitations

A limitation of the present study was that only one test-retest timespan was assessed (four weeks). Thus, the stability of retraction sensitivity and CIE susceptibility should be tested over longer time periods in future research. Doing this will allow us to gain a more acute understanding of the constructs' stability. Furthermore, as discussed, the current study investigated stability without manipulation of other cognitive and contextual factors, requiring such investigations to be done in future research. Moreover, while our observed-score correlation results provided evidence for *some* stability, the level of stability was relatively small (22% of variance). However, given the significantly higher level of stability shown in our CFAs, and the identified issues with internal-consistency reliability of the conventional CIE task, we recommend that future research assess internal-consistency reliability more thoroughly than previous experimental work and investigate plausible ways of improving the psychometric properties of the conventional CIE task; this would be particularly beneficial for studies interested in, or relying on, the test-retest reliability of said task. Finally, the conventional CIE task used in the current study provided, strictly speaking, a measure of retraction sensitivity, which is only one (arguably important) aspect of CIE susceptibility. While we collected data using an alternative CIE task that focused on misinformation reliance more specifically, the task failed to produce reliable measurement (see the Supplement for details). Therefore, future research should aim to address issues with the alternative task and attempt further test-retest investigations, as this may help to improve overall understanding of the CIE's temporal stability

Summary and Conclusion

In summary, the current study has demonstrated that retraction sensitivity, and thus potentially CIE susceptibility, has a detectable and significant level of stability over a four-week period. This corroborates past individual-differences research on the CIE, and is encouraging for this line of investigation in future research. These findings support the development of real-world CIE interventions that focus on improving information literacy for individuals with susceptibility to the CIE that is based in cognitive traits.

Acknowledgements

PM's contribution to this research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Declaration of Conflicting Interests

None of the authors have financial or personal conflicts of interest to disclose.

Funding

This research was supported by funding from the Australian Research Council (grant FT190100708, awarded to UKHE).

References

Arbuckle, J. L. (2019). *IBM SPSS Amos 26 User's Guide*. New York: IBM.

Baadte, C., & Dutke, S. (2013). Learning about persons: The effects of text structure and executive capacity on conceptual change. *European Journal of Psychology of Education*, *28*(3), 1045-1064. https://doi.org/10.1007/s10212-012-0153-2

Basol, M., Roozenbeek, J., & van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition*, *3*(1), 2. https://doi.org/10.5334/joc.91

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with non-normal data: Comparison of Pearson, Spearman, transformation, and resampling approaches. *Psychological Methods*, *17*(3), 399-417. https://doi.org/10.1037/a0028087

Blair, I. V., Lenton, A. P., & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review*, *9*(3), 590-596. https://doi.org/10.3758/BF03196317

Bonetto, E., Troïan, J., Varet, F., Lo Monaco, G., & Girandola, F. (2018). Priming resistance to persuasion decreases adherence to conspiracy theories. *Social Influence*, *13*(3), 125-136. https://doi.org/10.1080/15534510.2018.1471415

Boomsma, A. & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit & D. Sörbom (Eds.), *Structural equation modeling: Present and future. A Festschrift in honor of Karl Jöreskog* (pp.139-168). Scientific Software International. http://aboomsma.webhosting.rug.nl/csadata/boomsma_hoogland_2001.pdf

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

Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. *Model selection and multimodel inference, 2nd ed.* Springer.

Caruso, J. C. (2004). A comparison of the reliabilities of four types of difference scores for five cognitive assessment batteries. *European Journal of Psychological Assessment*, *20*(3), 166-171. https://doi.org/10.1027/1015-5759.20.3.166

Chang, E. P., Ecker, U. K. H., & Page, A. C. (2019). Not wallowing in misery—Retractions of negative misinformation are effective in depressive rumination. *Cognition and Emotion*, *33*(5), 991-1005. https://doi.org/10.1080/02699931.2018.1533808

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

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., Kawata, A., Kovvuri, A., Martin, J., Morgan, E., Sandhu, M., Sang, R., Scholz-Bright, R., Welch, A. T., Wolff, A. G., Zhou, A., & Nyhan, B. (2019). Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Political Behavior*, *42*, 1073-1095. https://doi.org/10.1007/s11109-019-09533-0

Cook, J. (2017). Understanding and countering climate science denial. *Journal & Proceedings of the Royal Society of New South Wales*, *150*(465-466), 207-219.

Cook, J., Bedford, D., & Mandia, S. (2014). Raising climate literacy through addressing misinformation: Case studies in agnotology-based learning. *Journal of Geoscience Education*, *62*(3), 296-306. https://doi.org/10.5408/13-071.1

Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, *12*(5), e0175799. https://doi.org/10.1371/journal.pone.0175799

Cronbach, L. J., & Furby, L. (1970). How we should measure “change’: Or should we? *Psychological Bulletin*, *74*(1), 68-80. https://doi.org/10.1037/h0029382

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

Ecker, U. K. H., & Ang, L. C. (2019). Political attitudes and the processing of misinformation corrections. *Political Psychology*, *40*(2), 241-260. https://doi.org/10.1111/pops.12494

Ecker, U. K. H., & Antonio, L. M. (2020). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. https://doi.org/10.31234/osf.io/qt4w8

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185-192. https://doi.org/10.1016/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*(2), 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*(4), 323-335. http://dx.doi.org/10.1037/xap0000028

Ecker, U. K. H., Lewandowsky, S., Fenton, O., & Martin, K. (2014). Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation. *Memory & Cognition*, *42*(2), 292-304. https://doi.org/10.3758/s13421-013-0358-x

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*(3), 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*(8), 1087-1100. https://doi.org/10.3758/MC.38.8.1087

Ecker, U. K., Sze, B. K., & Andreotta, M. (2021). Corrections of political misinformation: no evidence for an effect of partisan worldview in a US convenience sample. *Philosophical Transactions of the Royal Society B, 376*(1822), 20200145. https://doi.org/10.1098/rstb.2020.0145

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472-5477. https://doi.org/10.1073/pnas.1818430116

Forgas, J. P., & East, R. (2008). On being happy and gullible: Mood effects on skepticism and the detection of deception. *Journal of Experimental Social Psychology*, *44*(5), 1362-1367. https://doi.org/10.1016/j.jesp.2008.04.010

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502-1505. https://doi.org/10.1126/science.1255484

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences*, *42*(1), 37-48. https://doi.org/10.1016/j.paid.2006.06.019

Gignac, G. E. (2014). On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scales. *European Journal of Psychological Assessment*, *30*(2), 130-139. https://doi.org/10.1027/1015-5759/a000181

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

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201-209. https://doi.org/10.1016/j.jarmac.2013.10.001

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Heyman, T., Bruninx, A., Hutchison, K. A., & Storms, G. (2018). The (un) reliability of item-level semantic priming effects. *Behavior Research Methods*, *50*(6), 2173-2183. https://doi.org/10.3758/s13428-018-1040-9

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*(2), 305-324. https://doi.org/10.3758/s13421-013-0359-9

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82*(400), 1147-1149. https://doi.org/10.2307/2289392

Hotez, P. J. (2016). Texas and its measles epidemics. *PLoS Medicine*, *13*(10), e1002153. https://doi.org/10.1371/journal.pmed.1002153

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. http://dx.doi.org/10.1037/0278-7393.20.6.1420

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, *34*(1), 111-117. https://doi.org/10.1177/001316447403400115

Kendeou, P., Walsh, E. K., Smith, E. R., & O'Brien, E. J. (2014). Knowledge revision processes in refutation texts. *Discourse Processes, 51*(5-6), 374-397. https://doi.org/10.1080/0163853X.2014.913961

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E., Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., & Zaragoza, M. S. (2020). *The Debunking Handbook 2020*. https://sks.to/db2020. https://doi.org/10.17910/b7.1182

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, *13*(3), 106-131. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq War 2003. *Psychological Science*, *16*(3), 190-195. https://doi.org/10.1111/j.0956-7976.2005.00802.x

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433-442. https://doi.org/10.3758/s13428-016-0727-z

Lombardi, D., Danielson, R. W., & Young, N. (2016). A plausible connection: Models examining the relations between evaluation, plausibility, and the refutation text effect. *Learning and Instruction*, *44*, 74-86. https://doi.org/10.1016/j.learninstruc.2016.03.003

Maertens, R., Anseel, F., & van der Linden, S. (2020). Combatting climate change misinformation: Evidence for longevity of inoculation and consensus messaging effects. *Journal of Environmental Psychology*, *70*, 101455. https://doi.org/10.1016/j.jenvp.2020.101455

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior, 32*(2), 303-330. https://doi.org/10.1007/s11109-010-9112-2

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463-466. https://doi.org/10.1126/science.1183944

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163. https://doi.org/10.2307/271063

Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition*, *42*(1), 11-26. https://doi.org/10.3758/s13421-013-0339-0

Rapp, D. N., & Kendeou, P. (2007). Revising what readers know: Updating text representations during narrative comprehension. *Memory & Cognition*, *35*(8), 2019-2032. https://doi.org/10.3758/BF03192934

Rich, P. R., & Zaragoza, M. S. (2016). The continued influence of implied and explicitly stated misinformation in news reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 62-74. https://doi.org/10.1037/xlm0000155

Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570-580. https://doi.org/10.1080/13669877.2018.1443491

Sanderson, J. A., Gignac, G. E., & Ecker, U. K. (2021). Working memory capacity, removal efficiency and event specific memory as predictors of misinformation reliance. *Journal of Cognitive Psychology*, 1-15. https://doi.org/10.1080/20445911.2021.1931243

Schweizer, K. (2010). Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment*, *26*(1), 1-2. https://doi.org/10.1027/1015-5759/a000001

Seifert, C. M. (2002). The continued influence of misinformation in memory: What makes a correction effective? *Psychology of Learning and Motivation*, *41*, 265-294. https://doi.org/10.1016/S0079-7421(02)80009-3

Singh, K. A., Gignac, G. E., Brydges, C. R., & Ecker, U. K. H. (2018). Working memory capacity mediates the relationship between removal and fluid intelligence. *Journal of Memory and Language*, *101*, 18-36. https://doi.org/10.1016/j.jml.2018.03.002

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286-299. https://doi.org/10.1016/j.jarmac.2020.06.006

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

Swire-Thompson, B., 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*(1), 21-34. https://doi.org/10.1111/pops.12586

Taylor, L. E., Swerdfeger, A. L., & Eslick, G. D. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine*, *32*(29), 3623-3629. https://doi.org/10.1016/j.vaccine.2014.04.085

Thomas, D. R., & Zumbo, B. D. (2012). Difference scores from the point of view of reliability and repeated-measures ANOVA: In defense of difference scores for data analysis. *Educational and Psychological Measurement*, *72*(1), 37-43. https://doi.org/10.1177/0013164411409929

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, *112*(4), 578-598. https://doi.org/10.1037/0021-843X.112.4.578

Trafimow, D. (2015). A defense against the alleged unreliability of difference scores. *Cogent Mathematics*, *2*(1), 1064626. https://doi.org/10.1080/23311835.2015.1064626

Trevors, G., & Kendeou, P. (2020). The effects of positive and negative emotional text content on knowledge revision. *Quarterly Journal of Experimental Psychology*, *73*(9), 1326-1339. https://doi.org/10.1177/1747021820913816

Tsfati, Y., Boomgaarden, H. G., Strömbäck, J., Vliegenthart, R., Damstra, A., & Lindgren, E. (2020). Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, *44*(2), 157-173. https://doi.org/10.1080/23808985.2020.1759443

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wang, Y., McKee, M., Torbica, A., & Stuckler, D. (2019). Systematic literature review on the spread of health-related misinformation on social media. *Social Science & Medicine*, *240*, 112552. https://doi.org/10.1016/j.socscimed.2019.112552

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

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

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. *Quarterly Journal of Experimental Psychology*, *40*(2), 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*(1), 135-163. https://doi.org/10.1007/s11109-018-9443-y