

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Reasonable Forgetting: How Memory Shapes Belief Updating Over Time

Permalink

<https://escholarship.org/uc/item/3bk196pd>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Sanna, Greta Arancia

Lagnado, David

Ecker, Ullrich

et al.

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Reasonable Forgetting: How Memory Shapes Belief Updating Over Time

Greta Arancia Sanna (greta.sanna.23@ucl.ac.uk)¹, David Lagnado (d.lagnado@ucl.ac.uk)¹,
Ullrich Ecker (ullrich.ecker@uwa.edu.au)² &

Briony Swire-Thompson (b.swire-thompson@northeastern.edu)³

¹Department of Experimental Psychology, University College London

²School of Psychological Science, University of Western Australia

³Psychology of Misinformation Lab, Northeastern University

Abstract

Individuals often continue to rely on information even after credible sources have deemed it untrue, a phenomenon called the continued influence effect. This effect is especially consequential in politicised contexts where misinformation and polarisation threaten informed decision making. Existing accounts disagree on whether such belief persistence reflects cognitive failure or motivated reasoning, or whether it can instead arise from rational belief updating under constraints. We investigate this question by introducing delays during which no new information is provided, allowing memory for prior evidence to be less readily available. Across conditions, belief updating tracked the availability of evidence in memory, with beliefs regressing as supporting information decreased in availability and showing no systematic effects of political alignment. Moreover, belief trajectories aligned with predictions from Bayesian Network models parameterised using participants' own judgements. These findings suggest belief updating tracks memory availability and remains consistent with rational, evidence-based principles.

Keywords: Belief updating; Continued Influence Effect; memory availability; rational inference; Bayesian updating

Introduction

Effective belief updating is essential for navigating contemporary information environments in which evidence is uncertain, contradictory, and frequently revised. As new information emerges, individuals are required to revise beliefs that may have been formed on incomplete or erroneous evidence. Political misinformation is particularly well-suited to examine this challenge, as claims are often rapidly contested, corrected, and reframed, yet may continue to shape judgment and decision making long after their retraction (Walter & Tukachinsky, 2020). When belief updating fails, individuals may persist in relying on information that has been explicitly corrected, resulting in compromised judgment and suboptimal decisions. The current paper aims to examine the cognitive mechanisms underlying such apparent inconsistencies of belief updating, with a particular focus on how individuals respond to corrections in the face of competing and revisable information.

This persistence of misinformation following correction is commonly referred to as the Continued Influence Effect (CIE; Anderson et al., 1980; Guillory and Geraci, 2010; Johnson and Seifert, 1994; Ross et al., 1975; Wilkes and Leatherbarrow, 1988; Wilkes and Reynolds, 1999). Many early accounts of the CIE have implied that it could be an irrational process, often attributed to people failing to update their mental

models. Classic accounts propose that corrections disrupt an existing causal representation, leading individuals to revert to the original misinformation to resolve the resulting explanatory gaps (Brydges et al., 2018; Ecker et al., 2017; Gordon et al., 2017; Johnson & Seifert, 1994; Wilkes & Leatherbarrow, 1988). Related selective retrieval accounts suggest that corrected and uncorrected information coexist in memory, but are insufficiently discriminated at retrieval, allowing misinformation to continue influencing judgements (Ayers & Reder, 1998; Gordon et al., 2019; Kendeou et al., 2014). Within these frameworks, the CIE is often characterised as reflecting a deficit in belief revision processes, which some interpret as a cognitive failure (McIlhiney et al., 2023).

In political contexts, this interpretation has been extended to motivational accounts, where continued reliance on misinformation is attributed to identity protection, partisan bias, or motivated reasoning (Berinsky, 2017; Jost, 2017; Kahan, 2013; Swire-Thompson et al., 2020; Thaler, 2024). Some empirical findings show that corrections are less effective when they conflict with prior political worldviews, and this has been taken as evidence that the CIE reflects motivated distortion rather than epistemic evaluation (Druckman & McGrath, 2019; Nyhan & Reifler, 2010; Taber & Lodge, 2006). Under this account, politically congruent misinformation will be selectively retained, while incongruent information will become less readily available more rapidly. Other studies however have failed to replicate these findings (Ecker et al., 2021; Nyhan & Reifler, 2019).

In light of past research, the CIE remains theoretically ambiguous: observing that beliefs persist after correction does not, by itself, reveal whether this persistence reflects motivated distortion or rational updating given limited information, memory accessibility, or uncertainty about the correction. Recent work has challenged the assumption that the CIE necessarily reflects irrationality, arguing instead that belief updating may be broadly rational once factors such as how reliable people view the source, uncertainty, and evidential strength are taken into account (Connor Desai et al., 2020; Gershman, 2019; Haselton et al., 2009; Pilgrim et al., 2024). In particular, if a correcting source is perceived as unreliable, discounting the correction may be epistemically justified rather than motivated (Jern et al., 2014). Consistent with this account, Sanna et al. (2025) demonstrated that individuals are more likely to maintain belief in a corrected claim when the original source is perceived as more reliable than the

corrector. In such cases, continued influence reflects rational source evaluation rather than bias, aligning with evidence that disbelief in the correction is a key driver of the CIE (Ecker & Antonio, 2021).

Crucially, experiments typically study the continued influence effects in one study session, rather than over time. In real informational environments, however, corrective information is often encountered after substantial delays, and prior details may become increasingly difficult to retrieve over time. Memory therefore represents a critical but underexamined constraint on belief updating. Recent evidence suggests that belief revision may not be sustained over time, as corrective information is forgotten (Swire-Thompson et al., 2023). This phenomenon is known as belief regression, where people appear to re-endorse corrected information, and could be explained through opposing accounts.

Motivated reasoning accounts predict asymmetric regression, such that worldview-congruent misinformation is retained longer and exerts a more persistent influence than incongruent misinformation. By contrast, a rational account with memory constraints predicts that beliefs should track the availability and diagnosticity of evidence. This would cause beliefs to regress over time as supporting evidence is forgotten, with the effect occurring symmetrically across political alignments. That is, belief regression ought to occur whether the misinformation is congruent or incongruent with an individual's political affiliation. These accounts therefore make qualitatively distinct predictions about the pattern of belief and memory accessibility, allowing them to be empirically adjudicated.

To formalise these predictions, we adopt a Bayesian network framework for modelling belief updating in the presence of uncertain evidence from multiple sources. Within this framework, forgetting can be conceptualised as an inferential constraint that removes evidence from the network, thereby altering posterior beliefs without invoking motivational bias. Importantly, the Bayesian model is not intended to claim optimal or fully normative reasoning, but to generate qualitative predictions about belief trajectories over time. Under this account, belief updating should be independent of political alignment, and beliefs should remain stable following correction so long as corrective evidence remains accessible. The present study tests these predictions by examining belief updating and memory availability in political misinformation after it has been corrected.

The Study

Participants

We recruited 600 US participants from Prolific, half Republican half Democrats, and compensated them for their time (335 identified as female, 258 as male, 1 as non-binary and 4 as prefer to self-describe; $M_{\text{age}} = 47$, range = 19–81). Participants were compensated at a rate of \$9 per hour for an eight-minute initial study and \$18 per hour for a four-minute follow-up study. The target sample size was determined a

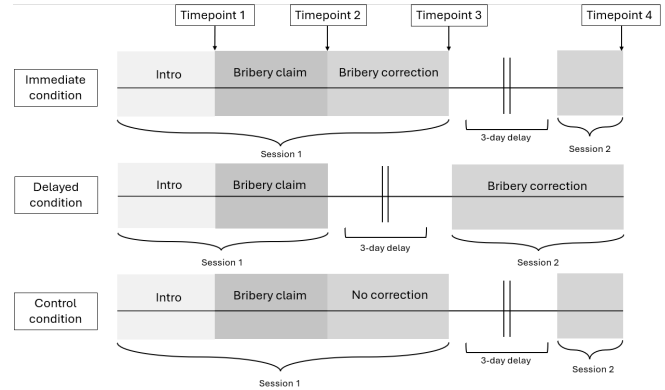


Figure 1: Experimental timeline for each condition. Belief in the bribery allegation was measured at each timepoint, while the three day delay was held constant across conditions.

priori to ensure sensitivity to small interaction effects involving political alignment, which are typical in misinformation correction research (e.g., $\eta_p^2 \approx .02$). A power analysis for the within-between interaction in a mixed-design ANOVA with 3 (correction timing: immediate vs. delay vs. control) $\times$ 2 (political affiliation of the politician: in-group vs. out-group) between-subjects conditions and 4 repeated measurements indicated that 186 participants would be sufficient to detect an effect of $f = .14$ with $\alpha = .05$ and power = .95. We conservatively recruited 600 to account for attrition, delayed measurement, and exploratory analyses. In line with pre-registered criteria, no participants were excluded. The study was pre-registered with supplementary materials available on OSF.

Design

This study used a 3 (correction timing: immediate vs. delay vs. control) $\times$ 2 (political affiliation of the politician: in-group vs. out-group) $\times$ 4 (timepoint) mixed factorial design, with timepoint as a within-subjects factor.

Procedure

Participants completed a University College London approved informed consent form before taking part. The study was administered online and consisted of two sessions separated by a three day delay.

Session 1 Session 1 included three belief measures: baseline (prior to the claim), post claim, and post correction (immediate condition only), see Figure 1. Participants first rated the credibility of two fictional news outlets, the *North Valley Times* and the *Independent Gazette*, using a six item adaptation of the Ohanian credibility scale (Ohanian, 1990).

They then read a vignette (adapted from Guillory & Geraci, 2013) introducing a fictional politician, Henry Light, whose political affiliation (Democrat or Republican) was randomly assigned. They were also introduced to the two fictional news

sources, and instructed that both were "considered to provide highly accurate reporting". Reading was self paced, although participants could not return to earlier sections once they had progressed. Midway through the vignette, all participants were informed that the *North Valley Times* reported that Henry Light had taken a bribe, after which they rated their belief in the bribery claim. Participants in the delay condition ended the first session there. Participants assigned to the immediate correction condition continued to read that the *Independent Gazette* issued a correction denying the claim. Participants in the no correction control condition received no correction. Participants then again rated their belief in the bribery claim and re-rated the credibility of the *North Valley Times*. Finally, they completed attention checks and a demographic questionnaire (age, gender, education, political affiliation).

Session 2 Participants were invited to return three days later. Although shorter than delays used in prior misinformation studies (e.g., Swire-Thompson et al., 2023), this interval was selected to account for the lower engagement or personal relevance of fictional politicians, which may accelerate forgetting and attenuate detectable effects. At the start of Session 2, all participants completed a free recall task about everything they could remember from Session 1. They then completed a delayed belief rating (post delay). For participants assigned to the delayed correction condition, the correction denying the bribery claim was presented following this belief judgment and an additional belief rating was asked following it. All participants then re-rated the credibility of the outlet that issued the correction. Session 2 therefore included two belief measures: post delay and post correction (in the delayed correction condition only).

Following the belief measures, participants completed a set of conditional probability judgments to elicit their assumptions about source reliability and claim generation. For both the *North Valley Times* and the *Independent Gazette*, participants judged the probability that the outlet would report the bribery claim under each combination of claim truth (true or false) and outlet reliability (reliable or unreliable). Reliability was defined as an outlet's perceived ability and willingness to provide accurate information. All conditional probability judgments were presented on a single page to allow joint consideration (see Supplementary Materials for details). These judgments were used to construct individualised conditional probability tables for Claim1 and Claim2 which then parameterised a personalised Bayesian network, enabling comparison between participants' observed belief updating and the predictions generated by the model, as well as assessment of how correction timing influenced belief revision.

Hypotheses

Our pre-registered hypotheses were:

H1: Participants will report lower belief in the bribery allegation when it concerns an ingroup politician than when it concerns an outgroup politician.

H2: From a rational account perspective, we hypothesize

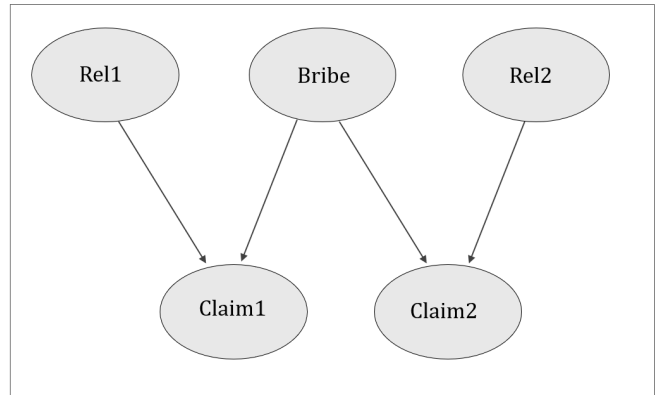


Figure 2: Bayesian Network of the political bribe vignette used in our study. Bribe = whether or not the politician took a bribe; Claim1 = news outlet accuses politician of taking a bribe; Claim2 = another news outlet corrects this claim; Rel1 = reliability of the news outlet accusing the politician of taking a bribe; Rel2 = reliability of news outlet correcting this claim.

that immediately following a correction, belief in the bribery allegation will return to baseline levels, both when the correction is presented immediately after the accusation (TP2) and when it is presented after a delay (TP3).

H3: We posit that in the immediate correction condition, the effect of the correction will diminish over time, such that belief in the allegation increases as the corrective information is forgotten (Swire-Thompson et al., 2023).

H4: In line with past work (e.g. Sanna & Lagnado, 2025), we predict that belief updating will align with Bayesian predictions such that changes in belief reflect the diagnosticity and availability of the evidence. This will occur similarly for both (a) political ingroup and (b) political outgroup members.

Bayesian Network model

To compare people's belief updating against a normative standard we use a Bayesian Network framework (Pearl, 1988). Bayesian networks (BNs) consist of two core components: (1) a directed acyclic graph encoding the probabilistic dependencies between variables, and (2) conditional probability tables associated with each variable that specify the strength of these dependencies and support quantitative belief updating via Bayes' rule. To capture the role of source reliability, we build on a Bayesian network framework originally proposed by Bovens and Hartmann (2003), which has also been applied to modelling witness reliability in legal contexts (e.g. Fenton et al., 2013; Lagnado, 2021). The Bayesian network corresponding to the vignette used in our study is illustrated in Figure 2. Binary variables are used to represent the central elements of the vignette: whether the politician accepted a bribe (*Bribe*); the allegation made by the accuser (*Claim1*); the correction issued by the corrector (*Claim2*); the reliability of the accuser (*Rel1*); and the reliability of the corrector

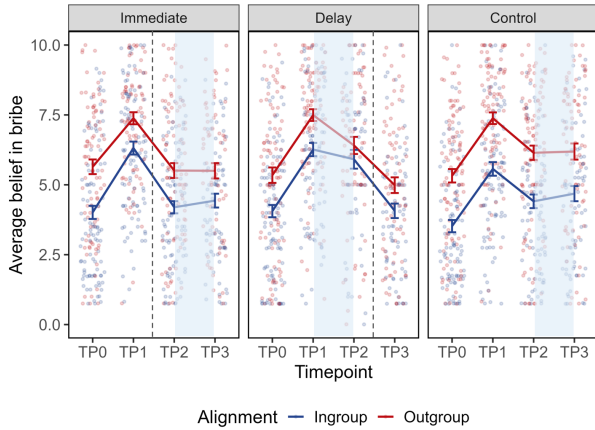


Figure 3: Belief in bribery across timepoints by political alignment. The shaded area indicates where the 3-day delay took place while the dotted line indicates when the correction was delivered. Note. Ratings ranged from 0 (very unlikely) to 10 (very likely) with error bars representing ± 1 standard error of the mean. Individual data points have been jittered along the x-axis to improve clarity and visibility.

(*Rel2*). Each claim is modelled as a function of both the state of the bribe (present or absent) and the reliability of the source issuing the claim (reliable or unreliable). The prior probabilities for *Bribe*, *Rel1*, and *Rel2*, as well as the conditional probability tables governing *Claim1* and *Claim2*, are derived from participants' individual judgments. Parameterising the network with these values allows us to compare participants' observed belief updating in response to the claims with the normative predictions generated by the Bayesian network.

Results

Political affiliation

Even before hearing the accusation, participants thought an ingroup politician ($M = 3.75$, $SE = 0.16$) would be less likely than an outgroup politician to take the bribe ($M = 5.31$, $SE = 0.16$), $F(1, 596) = 62.20$, $p < .001$, $\eta_p^2 = .094$.

To test whether participant-source congruence influenced the belief updating trajectory, we conducted a mixed ANOVA, with Timepoint (TP0–TP3) as a within-subject factor and *alignment* as a between-subject factor (Figure 3). There was a significant main effect of alignment indicating that ingroup participants consistently considered bribery to be less probable compared to outgroup participants (all $p < .001$). However, there was no interaction between alignment and timepoint ($p = .19$) suggesting the belief updating and subsequent belief regression unfolded similarly for the two groups. These findings support H1, indicating a stable distinction between ingroup versus outgroup belief in the bribe. This did not translate to asymmetric belief updating, with both groups showing similar belief updating.

Correction effectiveness

To evaluate whether the effectiveness of the correction differed between the immediate and delayed conditions, we ran a mixed ANOVA with *Timepoint* (baseline vs post-correction) as a within-subject factor and *correction timing* (immediate vs delayed) as a between-subject factor. The analysis showed no main effect of correction timing, $F(1, 308) = 2.20$, $p = .14$, and no correction timing by measurement timepoint interaction, $F(1, 308) = 0.89$, $p = .35$, indicating that the effectiveness of the correction did not vary depending on when it was delivered. Paired-sample *t*-tests further revealed that in the immediate condition belief after the correction did not significantly differ from baseline belief $t(190) = -0.21$, $p = .83$, $d = -0.02$, 95% CI $[-0.16, 0.13]$. Similarly, in the delayed condition there was no significant difference between post correction beliefs and prior beliefs $t(155) = 0.55$, $p = .59$, $d = 0.04$, 95% CI $[-0.11, 0.20]$. Bayesian analyses provided evidence in support of the null hypothesis in both cases ($BF_{01} = 12.05$ for the immediate condition; $BF_{01} = 10.00$ for the delayed condition). Collectively, these findings suggest, in line with H2, that the correction significantly returned participants back to baseline belief levels independent of timing.

Belief updating patterns

To test H3, we used a repeated-measures ANOVA comparing belief after the correction (TP2) to belief following the 3-day interval (TP3) in the immediate condition. As a reminder, motivated reasoning frameworks predict selective retention of worldview congruent information, whereas rational accounts predict symmetric belief updating independent of political alignment. The analysis revealed no significant main effect of measurement timepoint $F(1, 153) = 1.62$, $p = .21$, and a Bayesian analysis moderately supported the null hypothesis ($BF_{10} = 0.20$). These findings suggest that given belief had already returned to baseline levels following the correction, belief regression could not be observed in the subsequent timepoint.

To follow-up we examined belief change in the delay condition, where participants were not given any corrective information before the 3-day interval. A repeated-measures comparison of belief following the claim (TP1) and after the 3-day interval (TP2) revealed a significant decrease, $F(1, 155) = 12.72$, $p < .001$, with beliefs declining from $M_{TP1} = 6.88$ ($SE = 0.20$) to $M_{TP2} = 6.17$ ($SE = 0.23$). This suggests that as memory for the claim weakened, beliefs about the allegation drifted toward baseline levels. Crucially, the magnitude of this decline did not significantly differ between ingroup and outgroup participants, $F(1, 154) = 1.25$, $p = .27$, $\eta_p^2 = .01$. However, we note that a Bayesian independent-samples *t*-test provided only anecdotal support for the null hypothesis ($BF_{01} = 1.13$).

We next examined how many participants misremembered the political identity of the character. We found that over twice as many participants presented with an ingroup politician (26%) stated they remembered the bribery allegation in

the open text response than participants presented with an out-group politician (12%). A chi-squared test with Yates' continuity correction indicated an association between recall and political alignment, $\chi^2(1) = 5.71$, $p = .017$, which was corroborated by Fisher's exact test, $p = .013$, odds ratio = 0.39, 95% CI [0.17, 0.85]. Rather than suggesting a motivated reasoning account, these results suggest a relationship between the surprising nature of the evidence and recall, which affected subsequent belief. We conducted an exploratory analysis to test whether prior beliefs in the bribery allegation predicted later memory for the claim. The results revealed that participants who spontaneously recalled the bribery allegation during the delay had on average lower prior belief at baseline ($M = 30.97$, $SD = 25.28$, $n = 39$) than those who did not recall it ($M = 41.31$, $SD = 28.04$, $n = 166$). A logistic regression further supported this assumption with lower baseline beliefs significantly increasing the likelihood of recalling the allegation, $\beta = -0.015$, $SE = 0.007$, $z = -2.07$, $p = .039$. As such, memory for the claim appeared to be stronger when the allegation violated participants' initial expectations, consistent with a distinctiveness-based account of memory rather than motivated reasoning (Brown et al., 2007).

Among participants who did not explicitly state they recalled the allegation ($n=75$ out of 117), 64% showed decreased belief after the delay, potentially showing that when evidence was no longer accessible, conviction in the claim weakened. However, given the subgroup analysis is exploratory, this should be interpreted with caution.

In sum, the present findings provide no support for motivated reasoning accounts, which anticipate selective retention of worldview congruent information and as a result asymmetric belief updating. If anything, recall of the accusation was stronger when prior belief was lower, potentially suggesting higher retention for unexpected claims rather than motivationally congruent ones. Crucially, belief updating patterns over time did not differ by political alignment but instead tracked the availability of evidence to inform belief judgements. This pattern is consistent with a rational updating process in which beliefs weaken once the underlying evidence is no longer reliably retrieved.

Bayesian model comparison

For both conditions (see Figure 4 and 5), Bayesian-predicted posteriors were correlated with observed belief ratings across measurement timepoints (immediate: $r = .42-.59$; delayed: $r = .29-.56$, all $p < .001$). These findings indicate that participants' belief trajectories largely aligned with those predicted by the Bayesian model.

Deviation from Bayesian predictions showed a clear and systematic pattern across conditions. In the immediate correction condition, deviation did not vary reliably across measurement timepoints and was unaffected by political alignment (all $p > .10$). Although belief updating was initially more conservative than predicted following the allegation, deviation did not differ from zero after the correction and remained

non significant after the delay, indicating close alignment with the model once corrective evidence was introduced. In contrast, in the delayed correction condition, deviation varied by measurement timepoint, $F(2, 296) = 5.58$, $p = .009$, driven by increased conservatism during the correction delay interval, $t(148) = 3.42$, $p = .001$, likely reflecting memory-based attenuation of evidence weight not captured by the model. Crucially, political alignment did not influence deviation in either condition, suggesting that divergence from Bayesian predictions was driven by evidence availability rather than motivated distortion.

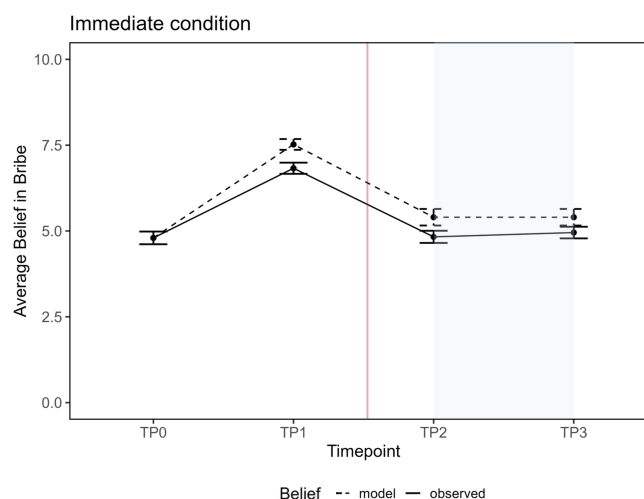


Figure 4: Immediate condition. Participants received the correction between TP1 and TP2 (red line), followed by a delay between TP2 and TP3 (shaded area). Ratings ranged from 0 (very unlikely) to 10 (very likely). Error bars represent ± 1 standard error of the mean.

Discussion

The current study examined belief updating in response to corrected misinformation over time, and evaluated whether the CIE is better explained by motivated reasoning or rational accounts that allow for forgetting. Even prior to the suggestion that the politician took a bribe, we found that ingroup members thought it was less likely than outgroup members (H1). However, subsequent belief updating remained consistent irrespective of whether a politician was an ingroup or outgroup member. Consequently, after the correction was delivered participants returned to their baseline belief regardless of political alignment or correction timing (H2). In other words, we found no continued influence effect. Notably, while beliefs post correction remained relatively high, comparison with baseline belief shows that belief ratings can be accounted for by a rational Bayesian process integrating priors and evidence diagnosticity.

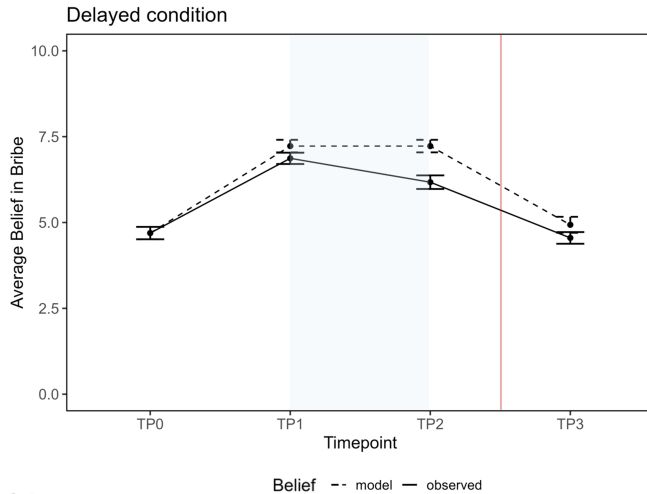


Figure 5: Delay condition. Participants experienced a delay between TP1 and TP2 (shaded area), followed by the correction between TP2 and TP3 (red line). Ratings ranged from 0 (very unlikely) to 10 (very likely). Error bars represent ± 1 standard error of the mean.

Forgetting over time In the immediate correction condition, given that belief returned to baseline after the correction was presented, belief regression could not be observed in the subsequent timepoint and as such beliefs remained stable over time (H3). One explanation for this pattern might be that when the accusation was remembered, the correction was also retained, and when the accusation was forgotten, the correction became redundant (Gilbert et al., 1993; Johnson & Seifert, 1999), though this remains to be directly tested. In either case, reduced memory for the correction during the delay would not be expected to produce systematic belief change. Crucially, there was no support for motivated reasoning whereby outgroup participants were more likely to discard corrective evidence than ingroup participants (Ecker et al., 2021; Nyhan et al., 2020).

In contrast, the delayed condition provided a critical test between motivated reasoning and forgetting based accounts of belief persistence. In this condition, delay periods led to participants forgetting the claim and, as a result, belief reduction. This suggests belief in the claim can be reduced simply due to forgetting processes.

These findings do not rule out motivated reasoning at earlier stages of the belief updating process (Sommer et al., 2023), such as evidence selection or source evaluation which can explain the reticence in altering political behaviour. Rational frameworks may also account for this justification; individuals might generate auxiliary hypotheses to dismiss evidence (Papakonstantinou et al., 2024) or hold their priors with such conviction that a single piece of evidence is insufficient to change their beliefs. Nonetheless, our evidence suggests that at the evidence evaluation stage, a motivated reasoning account is

not necessary to explain belief variation or persistence.

Rational vs motivated reasoning frameworks Our findings are more consistent with forgetting as the mechanism underlying belief updating than with motivated reasoning. While participants varied in their likelihood of recalling the claim, belief formation tracked the availability of recollected evidence, decreasing as the claim was forgotten rather than reflecting motivation to retain worldview congruent beliefs. Where motivated reasoning accounts (Jost, 2017; Kahan, 2013; Swire-Thompson et al., 2020) would predict asymmetrical belief updating in post-exposure intervals, our evidence favoured a rational framework in which evidence retrieval occurs independently of political alignment (Coppock, 2023). If anything, individuals appeared more likely to recall information that countered their prior beliefs (Frank et al., 2022; Greve et al., 2019). Where participants failed to remember the bribe, they tended to decrease their beliefs.

Bayesian Network comparisons further supported the forgetting account, with model predictions correlating with participants' belief updating trajectories (H4). Political alignment neither affected the predictive ability of the model nor predicted deviation from it, suggesting participants update their beliefs in line with reasonable assumptions even in politicised contexts under cognitive constraints. Memory thus appeared to function as a cognitive constraint: when evidence becomes unavailable, belief weakens accordingly. Apparent post correction belief differences may therefore arise from rational processes rather than motivated distortion.

Limitations and implications The present study has several limitations. In particular, the specificity of the vignette may hinder the generalisability of the findings. While participants clearly responded to the political nature of the vignette, as demonstrated by divergent priors, the fictional nature of the scenario may have attenuated the effects of political alignment. Additionally, requiring belief ratings after the delay may have prompted recollection of the vignette among participants, introducing a timing related design confound that limits generalisability.

Despite these constraints, the findings have practical implications for the design of misinformation corrections in real world settings. If belief persistence primarily reflects memory limitations rather than motivated resistance, corrective interventions should prioritise the durability and retrievability of accurate information (Ecker et al., 2011; Prike & Ecker, 2023). Corrections that are salient, repeatedly accessible, and well integrated with the original claim may therefore be more effective than those that merely challenge political identity. Notwithstanding these limitations, the present study supports rational accounts of belief updating, arguing that belief persistence in politicised contexts can arise from rational processes under memory constraints.

References

- Anderson, C., Lepper, M., & Ross, L. (1980). Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information. *Journal of Personality and Social Psychology*, 39, 1037–1049. <https://doi.org/10.1037/h0077720>
- Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, 5(1), 1–21. <https://doi.org/10.3758/BF03209454>
- Berinsky, A. J. (2017). Rumors and health care reform: Experiments in political misinformation. *British journal of political science*, 47(2), 241–262.
- Bovens, L., & Hartmann, S. (2003). Solving the Riddle of Coherence. *Mind*, 112(448), 601–633. <https://doi.org/10.1093/mind/112.448.601>
- Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological review*, 114(3), 539.
- 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 [Place: Netherlands]. *Intelligence*, 69, 117–122. <https://doi.org/10.1016/j.intell.2018.03.009>
- Connor Desai, S. A., Pilditch, T. D., & Madsen, J. K. (2020). The rational continued influence of misinformation. *Cognition*, 205, 104453. <https://doi.org/10.1016/j.cognition.2020.104453>
- Coppock, A. (2023). *Persuasion in parallel: How information changes minds about politics*. University of Chicago Press.
- Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change*, 9(2), 111–119. <https://doi.org/10.1038/s41558-018-0360-1>
- Ecker, U. K. H., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, 49(4), 631–644. <https://doi.org/10.3758/s13421-020-01129-y>
- Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, 6(2), 185–192. <https://doi.org/10.1037/h0101809>
- Ecker, U. K. H., Lewandowsky, S., 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., Sze, B. K. N., & 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: Biological Sciences*, 376(1822), 20200145. <https://doi.org/10.1098/rstb.2020.0145>
- Fenton, N., Neil, M., & Lagnado, D. A. (2013). A General Structure for Legal Arguments About Evidence Using Bayesian Networks [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12004>]. *Cognitive Science*, 37(1), 61–102. <https://doi.org/10.1111/cogs.12004>
- Frank, D., Kafkas, A., & Montaldi, D. (2022). Experiencing surprise: The temporal dynamics of its impact on memory. *Journal of Neuroscience*, 42(33), 6435–6444.
- Gershman, S. J. (2019). How to never be wrong. *Psychonomic Bulletin & Review*, 26(1), 13–28. <https://doi.org/10.3758/s13423-018-1488-8>
- Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of personality and social psychology*, 65(2), 221.
- Gordon, A., Brooks, J. C. W., Quadflieg, S., Ecker, U. K. H., & Lewandowsky, S. (2017). Exploring the neural substrates of misinformation processing [Place: Netherlands]. *Neuropsychologia*, 106, 216–224. <https://doi.org/10.1016/j.neuropsychologia.2017.10.003>
- Gordon, A., Ecker, U. K. H., & Lewandowsky, S. (2019). Polarity and attitude effects in the continued-influence paradigm [Place: Netherlands]. *Journal of Memory and Language*, 108. <https://doi.org/10.1016/j.jml.2019.104028>
- Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, 148(2), 325.
- Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences [Place: US]. *Psychonomic Bulletin & Review*, 17(1), 73–81. <https://doi.org/10.3758/PBR.17.1.73>
- Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, 2(4), 201–209. <https://doi.org/10.1016/j.jarmac.2013.10.001>
- Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias [Place: US]. *Social Cognition*, 27(5), 733–763. <https://doi.org/10.1521/soco.2009.27.5.733>
- Jern, A., Chang, K.-m. K., & Kemp, C. (2014). Belief polarization is not always irrational [Place: US]. *Psychological Review*, 121(2), 206–224. <https://doi.org/10.1037/a0035941>
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1420–1436. <https://doi.org/10.1037/0278-7393.20.6.1420>
- Johnson, H. M., & Seifert, C. M. (1999). Modifying mental representations: Comprehending corrections. *The construction of mental representations during reading*, 303–318.
- Jost, J. T. (2017). Ideological Asymmetries and the Essence of Political Psychology [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cogs.12004>].

- <https://onlinelibrary.wiley.com/doi/pdf/10.1111/pops.12407>. *Political Psychology*, 38(2), 167–208. <https://doi.org/10.1111/pops.12407>
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424. <https://doi.org/10.1017/S1930297500005271>
- 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>
- Lagnado, D. A. (2021, October). *Explaining the Evidence: How the Mind Investigates the World* [Google-Books-ID: OiJEEAAQBAJ]. Cambridge University Press.
- McIlhiney, P., Gignac, G. E., Ecker, U. K., Kennedy, B. L., & Weinborn, M. (2023). Executive function and the continued influence of misinformation: A latent-variable analysis. *Plos one*, 18(4), e0283951.
- Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking fact-checks literally but not seriously? the effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political behavior*, 42(3), 939–960.
- 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>
- Nyhan, B., & Reifler, J. (2019). The roles of information deficits and identity threat in the prevalence of misperceptions. *Journal of Elections, Public Opinion and Parties*, 29(2), 222–244.
- Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of advertising*, 19(3), 39–52.
- Papakonstantinou, T., Leong, K. I., & Lagnado, D. (2024). Humans generate auxiliary hypotheses to resolve conflicts in observational data. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 46.
- Pearl, J. (1988). Embracing causality in default reasoning. *Artificial Intelligence*, 35(2), 259–271.
- Pilgrim, C., Sanborn, A., Malthouse, E., & Hills, T. T. (2024). Confirmation bias emerges from an approximation to Bayesian reasoning. *Cognition*, 245, 105693. <https://doi.org/10.1016/j.cognition.2023.105693>
- Prike, T., & Ecker, U. K. (2023). Effective correction of misinformation. *Current Opinion in Psychology*, 54, 101712.
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm [Place: US]. *Journal of Personality and Social Psychology*, 32(5), 880–892. <https://doi.org/10.1037/0022-3514.32.5.880>
- Sanna, G. A., & Lagnado, D. (2025). Belief updating in the face of misinformation: The role of source reliability. *Cognition*, 258, 106090.
- Sanna, G. A., Pilditch, T. D., & Lagnado, D. (2025). The role of worldview congruence in misinformation correction: A bayesian approach to belief updating. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 47.
- Sommer, J., Musolino, J., & Hemmer, P. (2023). Updating, evidence evaluation, and operator availability: A theoretical framework for understanding belief. *Psychological Review*. <https://doi.org/10.1037/rev0000444>
- 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.
- Swire-Thompson, B., Dobbs, M., Thomas, A., & DeGutis, J. (2023). Memory failure predicts belief regression after the correction of misinformation. *Cognition*, 230, 105276. <https://doi.org/10.1016/j.cognition.2022.105276>
- Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3), 755–769.
- Thaler, M. (2024). The Fake News Effect: Experimentally Identifying Motivated Reasoning Using Trust in News. *American Economic Journal: Microeconomics*, 16(2), 1–38. <https://doi.org/10.1257/mic.20220146>
- 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. *The Quarterly Journal of Experimental Psychology Section A*, 40(2), 361–387. <https://doi.org/10.1080/02724988843000168>
- Wilkes, A. L., & Reynolds, D. J. (1999). On Certain Limitations Accompanying Readers' Interpretations of Corrections in Episodic Text. *The Quarterly Journal of Experimental Psychology Section A*, 52(1), 165–183. <https://doi.org/10.1080/713755808>