Belief Updating: Utilizing a New Paradigm to Separate Evidence Evaluation and Memory Dynamics

Gavin W. Oliver (olive467@umn.edu)[1] & Paul Schrater (schrater@umn.edu)[2]
[1]Department of Psychology, University of Minnesota
[2]Department of Computer Science & Engineering, University of Minnesota
[2]Department of Psychology, University of Minnesota

Abstract

People often judge claims using cues like familiarity and source credibility, but later judgments may also be shaped by memory for what they previously believed. We used a three-phase task to separate these influences. Prolific participants (N = 130) completed 30 pairs of factual statements across five domains. In Section 1, they rated belief in an initial statement (S1) and their familiarity with it. In Section 2, they rated belief in a conceptually related evidence statement (S2) paired with a source and judged source credibility and statement familiarity; S2 either supported or refuted S1. In Section 3, they reported what they thought they answered for S1 earlier (belief-memory) and then re-rated S1. Mixed-effects models showed that credibility and S2 familiarity predicted Section-2 belief, while belief-memory and final belief about S1 were associated with both initial belief and the implications of S2 evaluation.

Keywords: Belief updating; Source credibility; Memory; Mixed Effects Modeling

Introduction

The modern information environment gives people unprecedented access to knowledge, but it also confronts them with a constant stream of claims that are difficult or nearly impossible to directly verify. Because exhaustive fact-checking is rarely feasible, people rely on proxies for epistemic reliability—signals that are informative on average but imperfect in any specific case. Two major classes of such proxies are (a) content- and experience-based cues, such as plausibility, fluency, and familiarity, and (b) source-based cues, such as perceived expertise, trustworthiness, and credibility (Fricker, 2007). These shortcuts are adaptive in many everyday settings, yet they also create predictable vulnerabilities: misinformation can appear believable when it is easy to process, repeated, consistent with prior beliefs, or attributed to a seemingly credible source (Begg et al., 1992; Fazio et al., 2015; Lewandowsky et al., 2012; Swire-Thompson & Lazer, 2020; Unkelbach & Rom, 2017; Unkelbach, 2007).

A large body of empirical work shows that repetition and familiarity reliably increase perceived truth, even when repetition is unrelated to accuracy (the illusory truth effect; Bacon, 1979; Begg et al., 1992; Fazio et al., 2015; Dechêne et al., 2009). Mechanistically, repetition tends to increase processing fluency, and people may misinterpret this fluency as evidence of truth (Unkelbach, 2007; Unkelbach & Rom, 2017). In the present work, we distinguish between fluency—the ease of processing—and familiarity—the subjective sense of prior exposure. While fluency is often induced experimentally through repetition or perceptual manipulations, familiarity as measured here may reflect multiple underlying processes, including repetition, prior knowledge, and exposure. These effects are especially consequential because misinformation can continue to influence judgments even after corrections, a pattern documented in the continued influence and belief perseverance literatures (Lewandowsky et al., 2012; Ecker et al., 2010).

In parallel, persuasion research demonstrates that source credibility can shape belief, but not uniformly. Credibility effects depend on whether source information is treated as relevant and diagnostic, and how it interacts with other cues during judgment (Petty & Cacioppo, 1986; Priester & Petty, 1995; Kruglanski & Thompson, 1999; Pierro et al., 2004). Together, these findings suggest that belief formation reflects an interaction between information-based cues and source-based cues, filtered through prior beliefs and evaluative processes (Kunda, 1990; Sharot & Garrett, 2016; Ly et al., 2024).

Critically, however, belief updating is not only about judgment—it is also about memory. Source monitoring research shows that people do not reliably retain contextual details about where information came from; over time, source information may be forgotten even when content remains accessible, increasing the likelihood that familiarity drives later truth judgments (Johnson et al., 1993; Mitchell & Johnson, 2009; Conway et al., 1997). This creates a key pathway for delayed source effects, including classic accounts of the sleeper effect: the message persists while the discounting cue becomes less accessible (Gruder et al., 1978; Pratkanis et al., 1988). More broadly, belief change and belief persistence can arise from multiple mechanisms—some rooted in evaluation at the time of exposure (e.g., credibility-sensitive acceptance), and others rooted in later memory dynamics (e.g., source forgetting, misattribution, or reconstructive retrieval; Bartlett, 1932; Roediger & DeSoto, 2015). Recent work further suggests that memory failures can predict downstream belief regression and limit the effect of corrective information, underscoring the importance of measuring memory alongside belief updating (Wellons & Wahlheim, 2025).

A recent review of source credibility effects in misinformation research highlights a key methodological challenge: source credibility is often treated as a monolithic factor or confounded with other properties such as plausibility, familiarity, or ideological alignment, limiting

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

interpretability (Mang et al., 2024). Addressing this requires designs that (a) separately measure source-related dimensions, (b) manipulate whether new information supports or refutes prior beliefs, and (c) distinguish processes occurring at evaluation from those occurring at retrieval. The present study addresses these challenges using a three-phase design. Participants first report baseline graded beliefs (Section 1), then evaluate conceptually related statements paired with sources of varying perceived credibility (Section 2), and finally report both their memory for prior belief and their updated belief about the original statements (Section 3). Importantly, Section 2 statements are conceptually related rather than identical to the original claims, allowing responses to index evaluation of sourced evidence rather than simple repetition effects.

This structure is designed to dissociate how people evaluate new information from how they later remember and use prior beliefs. It also creates systematic opportunities for interaction among familiarity, credibility, and prior belief while reducing confounds common in real-world belief contexts. By separating these components, the present design allows us to identify when belief change reflects evaluation at exposure versus reconstruction at retrieval.

The present study focuses on three linked questions. First, how do familiarity and source credibility shape belief in new information, particularly when that information supports or refutes an existing belief? Second, how does exposure to new evidence influence memory for prior belief, and under what conditions does that memory reflect reconstructive bias? Third, how do these processes jointly determine later belief about the original claim? By addressing these questions within a unified framework, the present study aims to provide a more precise account of belief updating as the product of both evaluation and memory dynamics.

Methods

Participants

Participants were recruited via Prolific (U.S.; ages 18–60; English first language/fluency; no mental health issues; approval rate ≥ 99%). After providing informed consent, 149 participants completed some portion of the study; 19 were excluded, leaving **N = 130** for analysis (3,900 pair-level observations; 30 statement pairs per participant). Because analytic power calculations are not straightforward for crossed mixed-effects models, we conducted simulation-based sensitivity analyses for focal predictors (see Table 2 for focal predictors) using the homoscedastic final models; these are reported on OSF (https://osf.io/w9xzq/overview?view_only=e3dc953724ab433f92b9b533add9b93f). The average age was 43.4 years (SD = 9.9 years). The gender breakdown was 74 females, 48 males, and 8 didn't respond. Compensation was **$6.66** for ~40 minutes.

Exclusions. Participants were excluded for: failing an attention check (n = 8), incomplete data (n = 8), implausibly fast graded-belief responding (<1s on >50% of Section-1 trials; n = 2), or pervasive low-quality responding across multiple indicators (n = 1).

Materials and design

Stimuli were factual statements drawn from five knowledge domains (biology, history, physics, geography, psychology) and were designed to be politically neutral and noncontroversial (avoiding identity-relevant or ideologically charged claims). This constraint was used to minimize interpretive confounds that can dominate source judgments (e.g., motivated reasoning, partisan cueing) and to keep perceived "credibility" interpretable as an epistemic cue rather than a proxy for group alignment.

Stimulus construction proceeded in three steps. First, we generated an initial pool of candidate statements within each domain (assisted by large language model prompting) and then manually selected 30 statements (15 true, 15 false) that were clearly verifiable, unambiguous in meaning, and concise. Second, for each of these statements, we constructed a conceptually related partner statement that bore on the original claim without restating it verbatim. These partner statements were designed to either support or refute the conceptual content of the original statement (source_direction: support vs. refute), yielding 30 statement pairs (see Table 1 for example pairs). Critically, because S2 was conceptually linked to S1 rather than repeating it exactly, belief change observed at evidence evaluation is interpreted conservatively as updating with respect to the shared conceptual content of the pair, not necessarily literal revision of the specific S1 proposition. Direct re-evaluation of S1 occurred later during Section 3. The objective truth value of S2 was coded as truth_value (true vs. false).

Sources were preselected to be domain-appropriate to avoid implausible source–domain pairings that could trigger immediate source discounting. For each domain, six sources were selected (30 total; balanced between individuals and organizations), spanning a range of likely familiarity and perceived credibility. Sources were paired with statement pairs within the same domain.

Procedure

The task had three sections presented in a fixed order. For each section, participants completed the tasks of that section for all 30 statements before moving on to the next section. In Section 1, participants rated their belief in an initial statement (S1) and then rated S1 familiarity. In Section 2, participants evaluated a source (expertise and trust) and then rated belief and familiarity for a second, conceptually related statement (S2). In Section 3, participants saw S1 again, attempted to reproduce their earlier belief rating (belief-memory), and then provided a new belief rating for S1.

Table 1: Example stimulus structure

Section 1 (S1)	Section 2 (S2)	Relation	Source
"Poland and Hungary share a border with Germany."	"Poland and Hungary border each other."	Support	Bear Grylls
"The U.S. joined the Allies in World War II right after Germany invaded Poland."	"The United States entered World War II following the attack on Pearl Harbor."	Refute	History Channel

Measures

Belief (B1, B2, B3). Belief was measured with a 0–100 slider ("How likely do you think it is that this statement is true?"). 7 tick-mark labels were displayed as interpretive guides (e.g., "Certainly False" … "Unsure / equally likely true or false" … "Certainly True"), and participants were instructed to use the full scale rather than only endpoints/midpoint.

Statement familiarity (fam_S1, fam_S2). After each statement, participants answered "How familiar does this statement feel to you?" on a 5-point scale with anchors: (1) "Completely unfamiliar—'I've never seen or heard anything like it before'," (2) "Somewhat familiar—'It seems vaguely related to something I've heard'," (3) "Moderately familiar—'It feels familiar, but I'm not sure where from'," (4) "Very familiar—'I've seen or heard something very similar'," and (5) "Recognize—'I've definitely seen or heard this exact statement before'."

Perceived source credibility (expertise, trust). Participants then rated (a) source expertise ("How knowledgeable do you think this source is about [domain] topics?"; Not at all → Extremely) and (b) trust ("How likely are you to trust this source to provide accurate information about [domain] topics?"; not at all likely → extremely likely). Because expertise and trust were highly correlated, they were averaged into a credibility composite and z-scored for analysis.

Belief memory (mem_resp). Section 3, participants first completed a belief-memory prompt (e.g., "Please try to recall where on the 0–100 scale you clicked earlier… Now please try to click on the same spot on the scale below"), producing mem_resp.

Coding and preprocessing

Belief ratings (B1, B2, B3) and memory were centered to the interval [-1, 1] via $y=(x-50)/50$, yielding B1c, B2c, B3c, mem_resp_c. Familiarity ratings and perceived source credibility were z-scored (fam_S1_z, fam_S2_z, source_cred_z). We sum-coded truth_value (reference level = truth) and source_direction (reference level = support) and scaled to +-1 so main effects reflect averages across levels and interactions reflect deviations from those averages.

To represent the implication of the Section 2 judgment for the original Section 1 claim, we computed an implied evidential shift variable (evid_shift_c). Because Section 2 statements were conceptually related to, but not identical with, Section 1 statements, B2c was not treated as a direct updated judgment of the original claim. Instead, Section 2 judgments were aligned by source direction: for support trials, B2c was retained, whereas for refute trials, B2c was reverse-coded so that higher values always indicated stronger implied support for the original Section 1 claim. We then subtracted baseline belief (B1c) from this aligned Section 2 judgment. Thus, positive evid_shift_c values indicate that the Section 2 evaluation implied a more positive belief about the original claim than the participant's baseline response, whereas negative values indicate an implied shift toward lower belief. This variable was used in downstream models because raw B2c combines evaluation of the Section 2 statement with prior-belief carryover and differs in implication across support and refute trials. evid_shift_c places Section 2 evaluations on a common scale representing their implication for the original claim while preserving the distinction between evaluation of S2 and direct belief updating about S1.

Analytic strategy and robustness checks

We fit three linear mixed-effects models (lme4; maximum likelihood) predicting (i) belief in sourced evidence (B2c), (ii) memory for prior belief (mem_resp_c), and (iii) later belief about the original statement (B3c). Fixed effects were selected based on theoretical considerations and cross-validation comparisons, with full specifications reported on OSF. Random effects followed a maximal structure where supported by the data (Barr et al., 2013), with simplification only when required for convergence or stability. Estimated marginal means and trends were obtained using emmeans with Kenward–Roger degrees of freedom. Robustness checks included residual diagnostics, participant-level influence analyses, alternative variance structures, and Gaussian-transformed outcome re-estimation; full results are reported on OSF.

Results

Overview

We focus on three primary outcomes corresponding to distinct stages of belief processing: (1) evaluation of sourced evidence in Section 2 (B2c), (2) memory for prior belief in Section 3 (mem_resp_c), and (3) final belief about the original statement in Section 3 (B3c). All models were estimated using mixed-effects regression. Table 2 summarizes the dependent variables, psychological roles, and focal predictors for each model (see OSF for full specifications and robustness analyses).

endorsement was near zero, β = -0.006, SE = 0.049, 95% CI [-0.101, 0.089]. In the refute condition, stronger initial

Table 2: Main models specification

Model	Dependent variable	Psychological question	Focal fixed effects reported	Random structure	Residual structure
Evidence evaluation	B2c: belief in the sourced Section 2 statement	What predicts acceptance of new sourced evidence?	source_cred_z, fam_S2_z, B1_pos × source_direction, B1_neg × source_direction	Random intercepts/slopes for pair and participant; final structure selected by fit, convergence, and stability	Heteroskedastic by truth_value
Belief memory	mem_resp_c: memory for original Section 1 belief	Does later memory reflect both original belief and intervening evidence?	B1c, evid_shift_c, evid_shift_c × source_direction,	Random intercepts/slopes for participant and pair; final structure selected by fit, convergence, and stability	Full 2 × 2 cell-specific variance
Final belief	B3c: later belief about original Section 1 statement	Does final belief reflect original belief, remembered belief, and evidence shift?	B1c, mem_resp_c, evid_shift_c	Random intercepts/slopes for participant and only one random slope for pair; final structure selected by fit, convergence, and stability	Heteroskedastic by source_direction

endorsement was associated with lower belief in the refuting evidence, β = -0.265, SE = 0.047, 95% CI [-0.357, -0.173]. Thus, participants who more strongly believed the original claim were less likely to accept evidence that refuted it, consistent with discounting of belief-inconsistent information.

Taken together, these results suggest that evaluation of sourced evidence reflects both general cue-based influences (credibility and familiarity) and a conditional integration of prior belief that depends on whether the evidence is belief-consistent or belief-inconsistent relative to participants' initial beliefs.

Evidence Evaluation: Belief in Sourced Statements (B2c)

We first examined how participants evaluated the sourced evidence statement (S2) in Section 2. This measure reflects belief in a novel, source-attributed statement that either supports or refutes the original claim (S1), rather than a direct revision of the original belief.

Two cue-based predictors showed clear effects. Higher perceived source credibility was associated with greater belief in S2 (source_cred_z; β = .110, SE = .017, z = 6.63, p <.001), and higher familiarity with S2 also increased belief (fam_S2_z; β = .321, SE = .020, z = 16.04, p <.001). These results are consistent with prior work showing that both credibility and familiarity independently contribute to truth judgments.

For baseline disbelief (B1_neg), the relation between prior belief and belief in the sourced evidence depended on whether the evidence supported or refuted the original claim. In the support condition, stronger initial disbelief was associated with lower belief in the sourced evidence, β = -0.366, SE = 0.045, 95% CI [-0.454, -0.277]. In the refute condition, this slope was small and uncertain, β = 0.050, SE = 0.044, 95% CI [-0.036, 0.137]. Thus, when the new statement supported the original claim, participants who initially rejected the original claim were less likely to accept the supporting evidence; when the new statement refuted the original claim, initial disbelief did not strongly predict evaluation of the evidence.

For baseline endorsement (B1_pos), the pattern was complementary. In the support condition, the slope of prior

Belief Memory: Memory for Prior Belief (mem_resp_c)

We next examined participants' memory for their original belief about S1. This measure reflects reconstructive retrieval rather than simple recall, allowing us to test whether intervening evidence influences remembered belief. As expected, remembered belief strongly tracked the original belief: participants tended to recall their prior belief in line with what they initially reported (B1c; β = .730, SE = .122, z = 22.72, p <.001). However, memory was not determined solely by the original response.

Evidence from Section 2 also influenced remembered belief. Positive implied evidential shifts (evid_shift_c) were associated with more positive remembered prior beliefs, above and beyond participants' actual baseline beliefs (evid_shift_c; β = .122, SE = .020, z = 6.06, p <.001).

This effect depended on the direction of the evidence (β = .050, SE = .016, z = 3.09, p =.002). Implied evidential shifts predicted later memory for prior belief in both source-direction conditions, but the association was stronger when the evidence supported the original claim. In the support condition, larger implied shifts were associated with stronger shifts in remembered prior belief, $\beta = 0.184$, SE = 0.022, 95% CI [0.140, 0.228]. The same association was positive but smaller in the refute condition, $\beta = 0.086$, SE = 0.020, 95% CI [0.046, 0.126]. This suggests that intervening evidence shaped memory for prior belief, with supporting evidence more strongly incorporated into later memory than refuting evidence.

Overall, these results show that memory for prior belief is not a passive record but is reconstructed based on both original belief and intervening evidence.

Final Belief: Later Judgment of the Original Statement (B3c)

Finally, we examined participants' later belief about the original statement (S1). This measure captures belief after both exposure to new evidence and a memory retrieval step.

Final belief was influenced by three key components. First, original belief continued to predict later belief (β = .203, SE = .019, z = 10.53, p <.001), indicating persistence of prior belief over time. Second, remembered prior belief strongly predicted final belief (β = .758, SE = .020, z = 37.72, p <.001), suggesting that reconstructed memory serves as an important input to later judgments. Third, evidence-implied shifts from Section 2 continued to influence final belief even after accounting for both original belief and remembered belief (β = .112, SE = .019, z = 5.79, p <.001).

This pattern indicates that the influence of evidence on later belief is not fully mediated by memory. Instead, evidence appears to affect final belief through two pathways: indirectly, by altering memory for prior belief, and directly, by contributing additional influence at the time of later judgment.

Robustness of Findings

Across all three focal models, results were robust to multiple analytic checks, including residual diagnostics, influential-participant exclusion, alternative residual variance structures, and Gaussian-transformed outcome specifications. While some secondary effects varied across specifications, the primary conclusions reported here remained stable. Full robustness analyses and model specifications are available on OSF.

Discussion

The present study examined how source credibility, familiarity, and prior belief shape belief formation across three temporally separated stages: evaluation of new information, memory for prior belief, and later belief about the original claim. By separating these stages, the design allows belief-related processes that are typically conflated, evaluation of evidence and reconstruction of prior belief, to be examined independently. Across models, the results converge on a central pattern: information encountered after an initial judgment influences later belief through two partially separable pathways. One pathway operates indirectly, through changes in how prior belief is remembered, whereas a second pathway reflects the continued influence of previously evaluated information at the time of later judgment.

Evaluation of sourced statements (Section 2) reflected both cue-based influences and the relationship between new information and prior belief. Familiarity increased belief, consistent with extensive evidence that repeated or fluent information is more likely to be judged as true (Bacon, 1979; Dechêne et al., 2009; Fazio et al., 2015). Source credibility also increased belief, aligning with research showing that source characteristics influence acceptance when individuals attend to and utilize such cues (Petty & Cacioppo, 1986; Priester & Petty, 1995; Mang et al., 2024). The B2c results further suggest that prior beliefs are recruited asymmetrically during evidence evaluation. Initial belief strength predicted responses primarily when the Section 2 statement conflicted with the participant's prior belief, but had little influence when the statement was belief-consistent. This pattern aligns with contemporary accounts of motivated reasoning and biased assimilation, in which belief-inconsistent information is subjected to greater scrutiny, whereas belief-consistent information is accepted with relatively less evaluation (Taber & Lodge, 2006; Sharot & Garrett, 2016). One interpretation is that belief-inconsistent information triggers greater comparison against existing beliefs, leading participants to rely more strongly on prior belief when forming judgments. Nevertheless, the results suggest that prior belief does not uniformly influence evidence evaluation, but instead is selectively recruited under conditions of conflict.

Beyond evaluation, intervening information also influenced how prior beliefs were later remembered. Memory for prior belief (Section 3) reflected both baseline belief and the influence of information encountered after the initial judgment, indicating that recollection was reconstructive rather than a veridical record (Bartlett, 1932; Johnson et al., 1993; Roediger & DeSoto, 2015). To capture the contribution of intervening information, we computed an implied evidential shift variable (evid_shift_c), which indexes how participants' evaluations of Section 2 statements imply a change in belief about the original claim, relative to baseline belief.

Using this measure, we found that evid_shift_c predicted memory for prior belief above and beyond baseline belief, indicating that memory reports were systematically associated with the implications of intervening evaluations. Notably, this association was stronger for support than refute trials, suggesting that the implications of supporting statements were more readily reflected in reconstructed memory than those of refuting statements. Because this effect is defined by the structural relation between statements rather than participants' initial beliefs, it suggests that alignment between new information and the original claim facilitates integration into memory representations, whereas refuting information may be more likely to be maintained as distinct or conflicting information. Consistent with reconstructive accounts of memory, these results indicate that memory for prior belief reflects a combination of the original belief and subsequently encountered information, shaped in part by how that information relates to the original claim (Johnson et al., 1993; Mitchell & Johnson, 2009).

The final belief model further indicates that the influence of intervening information is not fully accounted for by memory. Even after controlling for both baseline belief and remembered prior belief, implied evidential shift continued to predict later belief. This suggests that the implications of Section 2 evaluations were not solely reflected in reconstructed memory, but also remained associated with final judgments. One interpretation is that information encountered after the initial judgment contributes to belief both through memory reconstruction and through additional processes at the time of evaluation or later judgment. This interpretation is consistent with recent work on memory and belief integration, which emphasizes that newly encountered information can influence both how prior beliefs are remembered and how subsequent judgments are formed (Wellons & Wahlheim, 2025). More broadly, the pattern aligns with research showing that previously encountered information can continue to influence reasoning even when it is not fully incorporated into explicit beliefs or memory reports (Lewandowsky et al., 2012; Ecker et al., 2010).

Taken together, these findings are consistent with a dual-pathway account of belief formation. One pathway operates through memory reconstruction: evaluations of new information are associated with how prior beliefs are remembered, and these reconstructed memories then relate to later judgments. A second pathway reflects that the implications of previously evaluated information remain associated with belief at the time of judgment, beyond what is captured by memory. This distinction helps reconcile findings across literatures that have emphasized reconstructive memory processes (Bartlett, 1932; Roediger & DeSoto, 2015) and the continued influence of information (Lewandowsky et al., 2012). The present results suggest that both processes may operate simultaneously and contribute to belief formation.

These findings also have implications for theoretical models of belief updating. Many models treat updating as a unitary process in which new information is integrated with prior belief at the time of judgment (Kruglanski & Thompson, 1999). The present results suggest that this view is incomplete. Instead, belief formation reflects at least two separable processes: evaluation of new information and reconstruction of prior belief. In addition, the results highlight the importance of distinguishing between prior belief and memory for prior belief. Because memory is reconstructive and influenced by intervening information, it cannot be treated as a direct proxy for prior belief. Models that do not separate these components may misattribute effects of memory reconstruction to belief updating or vice versa.

Several limitations should be noted. First, the relationship between Section 1 and Section 2 statements was defined at a structural level (support vs. refute), but the degree and nature of this relationship likely varied across items and participants. Because the statements were semantically related rather than identical or directly negated, participants may have interpreted the connection between statements in different ways, introducing variability in how intervening information was integrated with prior belief. This variability may have contributed additional noise to measures of evidential influence, including the implied evidential shift variable. Future work could address this limitation by using more tightly controlled stimulus pairs, such as identical statements, direct negations, or systematically manipulated levels of semantic overlap. Second, familiarity in the present study reflects subjective judgments that may capture multiple underlying processes, including fluency, repetition, and prior knowledge (Fazio et al., 2015; Unkelbach & Rom, 2017), limiting precision in interpretation. Finally, although the results revealed differences between support and refute conditions in how evaluation influenced memory, the present design did not fully test how these effects interact with the direction of initial belief. Future research could examine whether the influence of intervening information on memory varies depending on whether the original claim was initially endorsed or rejected.

By separating evaluation, memory, and later judgment, the present study provides a more precise account of how belief formation unfolds over time. Information influences belief both through memory-mediated reconstruction and through direct integration at judgment, and these processes are systematically shaped by the relationship between new information and prior belief. Understanding belief formation therefore requires attention not only to what information people encounter, but also to when and how that information is evaluated and remembered.

Acknowledgments

An AI language model (ChatGPT, OpenAI) was used as an assistive tool during manuscript preparation to help revise phrasing and improve clarity of sections written by the authors and to suggest alternative sentence-level formulations. It was also used during materials development to brainstorm candidate statements and supporting/refuting pairings across domains. All final text, experimental design decisions, stimulus selection, truth verification, analyses, and interpretations were made by the authors. The AI tool did not have access to the study data and was not used to compute or validate statistical results.

Supplemental materials

All supplemental materials are included on osf. OSF link: https://osf.io/w9xzq/overview?view_only=e3dc953724ab433f92b9b533add9b93f

References

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 241-252. https://doi.org/10.1037/0278-7393.5.3.241

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-273. https://doi.org/10.1016/j.jml.2012.11.001

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. *Cambridge University Press.*

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*(4), 446-458. https://doi.org/10.1037/0096-3445.121.4.446

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology. General, 126*(4), 393–413. https://doi.org/10.1037//0096-3445.126.4.393

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*(2), 238-257. https://doi.org/10.1177/1088868309352251

Ecker, U. K., Lewandowsky, S., & Tang, D. T. (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

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General, 144*(5), 993–1002. https://doi.org/10.1037/xge0000098

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. *Oxford University Press.*

Gruder, C. L., Cook, T. D., Hennigan, K. M., Flay, B. R., Alessis, C., & Halamaj, J. (1978). Empirical tests of the absolute sleeper effect predicted from the discounting cue hypothesis. *Journal of Personality and Social Psychology, 36*(10), 1061–1074. https://doi.org/10.1037/0022-3514.36.10.1061

Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition, 20*(4), 1705-1721. https://doi.org/10.1016/j.concog.2011.08.018

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28. https://doi.org/10.1037/0033-2909.114.1.3

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the Unimodel. *Psychological Inquiry, 10*(2), 83–109. https://doi.org/10.1207/S15327965PL100201

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. https://doi.org/10.1037/0033-2909.108.3.480

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

Ly, D. P., Bernstein, D. M., & Newman, E. J. (2024). An ongoing secondary task can reduce the illusory truth effect. *Frontiers in Psychology, 14*, 1215432. https://doi.org/10.3389/fpsyg.2023.1215432

Mang, V., Fennis, B. M., & Epstude, K. (2024). Source credibility effects in misinformation research: A review and primer. *Advances in Psychology,* aip00028. https://doi.org/10.56296/aip00028

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638–677. https://doi.org/10.1037/a0015849

Petty, R.E., Cacioppo, J.T. (1986). Message elaboration versus peripheral cues. In: Communication and Persuasion. *Springer Series in Social Psychology.* Springer, New York, NY . https://doi.org/10.1007/978-1-4612-4964-1_6

Pierro, A., Mannetti, L., Kruglanski, A. W., & Sleeth-Keppler, D. (2004). Relevance override: On the reduced impact of "cues" under high-motivation conditions of persuasion studies. *Journal of Personality and Social Psychology, 86*(2), 251–264. https://doi.org/10.1037/0022-3514.86.2.251

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper effect. *Journal of Personality and Social Psychology, 54*(2), 203–218. https://doi.org/10.1037/0022-3514.54.2.203

Priester, J. R., & Petty, R. E. (1995). Source attributions and persuasion: Perceived honesty as a determinant of message scrutiny. *Personality and Social Psychology Bulletin, 21*(6), 637-654. https://doi.org/10.1177/0146167295216010

Roediger, H. L., & DeSoto, K. A. (2015). Reconstructive memory, psychology of. *In International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (pp.50-55). Elsevier Inc.. https://doi.org/10.1016/B978-0-08-097086-8.51016-2

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. *Trends in Cognitive Sciences, 20*(1), 25-33. http://dx.doi.org/10.1016/j.tics.2015.11.002

Swire-Thompson, B., & Lazer, D. (2020). Public health and online misinformation: challenges and recommendations. *Annual Review of Public Health, 41*, 433-451. https://doi.org/10.1146/annurev-publhealth-040119-094127

Taber, C.S. & Lodge, M. (2006), Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3) 755-769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*(1), 22-38. https://doi.org/10.1016/j.concog.2008.09.006

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition, 160*, 110-126. https://doi.org/10.1016/j.cognition.2016.12.016

Wellons, B.M., Wahlheim, C.N. (2025). Misinformation reminders enhance belief updating and memory for corrections: the role of attention during encoding revealed by eye tracking. *Cognitive Research 10*, 39 https://doi.org/10.1186/s41235-025-00649-y