PNAS Nexus, 2024, **3**, pgae490
https://doi.org/10.1093/pnasnexus/pgae490
Advance access publication 1 November 2024
Research Report

Continued influence of false accusations in forming impressions of political candidates

Michael S. Cohen[a,b,*], Victoria Halewicz[a,c], Ece Yildirim[a] and Joseph W. Kable[a,*]

[a]Department of Psychology, University of Pennsylvania, 425 S. University Ave., Philadelphia, PA 19104, USA
[b]Department of Psychology, University of Chicago, 5848 S. University Ave., Chicago, IL 60637, USA
[c]Department of Cognitive, Linguistic & Psychological Sciences, Brown University, 190 Thayer St., Providence, RI 02912, USA
*To whom correspondence should be addressed: Email: mscohen@uchicago.edu (M.S.C.); Email: kable@psych.upenn.edu (J.W.K.)
Edited By David Rand

Abstract

Previous work has shown that false information affects decision-making even after being corrected, a phenomenon known as "continued influence effects" (CIEs). Using mock social media posts about fictional political candidates, we observe robust within-participant CIEs: candidates targeted by corrected accusations are rated more poorly than candidates not targeted by allegations. These effects occur both immediately and after as much as a 2-day delay. We further demonstrate that vulnerability to CIEs in a political context varies systematically between individuals. We found that certain groups are more susceptible to CIEs on immediate candidate ratings (i) those who rely more on intuitive feelings, (ii) those with lower digital literacy knowledge, and (iii) younger individuals. These individuals' judgments appear to be relatively more influenced by the refuted accusations and/or less influenced by the factual refutations. Interestingly, political orientation did not affect CIEs, despite its influence on explicitly identifying misinformation. Moreover, people recalled accusation stimuli better than refutations at a delay, suggesting that emotions may drive the prioritized processing of accusations. Our results indicate that analytic thinking could be protective when people judge political candidates targeted by refuted false information.

Keywords: continued influence effect, misinformation, analytic thinking, digital literacy, individual differences

Significance Statement

False information, even after being corrected, can still influence subjective impressions and decisions about its targets. We address this issue using a novel approach: presenting mock social media posts regarding a large set of fictional political candidates. Refuted misconduct allegations yield lower candidate ratings both immediately after the posts were presented and at 30-min and 2-day delays. Consistent individual differences are observed on the immediate ratings, as individuals who self-report relying more on intuitive thinking, who demonstrate lower digital literacy, or are younger, are more vulnerable to influence from corrected false information. Since real-world harms from misinformation occur primarily via distortion of decision-making, our approach elucidates who is more vulnerable to making poor decisions based on false information.

Introduction

Content that evokes strong moral and emotional responses tends to receive greater social media engagement (1). False information often elicits more negative emotions than true content (2), causing it to spread rapidly across social media platforms. In an ideal world, factual refutations would reverse the harmful effects of misinformation. In actuality, even when people learn of the falsehood, inaccurate information can still influence their later judgments, a phenomenon known as "continued influence effects" (CIEs) (3). Past work has documented some cognitive mechanisms as well as correction strategies for CIEs (4, 5). Still, CIEs typically persist even when best practices for corrections are followed, thus it remains crucial to understand the lingering impacts of false information on behavior. We measure CIEs in the context of accusations of political misconduct using novel stimuli designed to resemble social media postings, and examine demographic and psychometric predictors of the magnitude of these effects in political decision-making.

CIEs were first studied in the domain of causal reasoning (3), for example, about the causes of a fire. Studies in this domain have proposed potential cognitive mechanisms for CIEs. One possible mechanism is a failure to integrate new information into one's mental model (e.g. 3). An alternate mechanism is selective retrieval (5, 6)—automatically remembering misinformation but not remembering its refutation, and failing to engage in strategic retrieval sufficient to associate the two. Individual differences in

Competing Interest: The authors declare no competing interests.
Received: May 11, 2024. **Accepted:** October 10, 2024

CIE vulnerability have been linked to cognitive abilities like vocabulary knowledge (7), working memory capacity (8), and long-term memory for stimulus details (9). However, working memory's predictive power disappears when controlling for long-term memory (9). Although the present studies do not focus primarily on these mechanisms, we measure recognition memory in part to test a simple selective retrieval account of CIEs.

Political misinformation has gained significant societal attention in recent years, but there are challenges to integrating the cognitive psychology literature on CIEs with research on political misinformation. One key question is whether and when CIEs occur when forming impressions about other people. CIEs were observed in one study using hypothetical political candidates, in which refuted misconduct allegations still negatively impacted candidate evaluations (10). Classic social psychology research has shown that impressions of task ability, both for oneself and an observed other person, remain influenced by false feedback even after debriefing (11). Similarly, evidence presented but then declared inadmissible tends to impact real-world jury verdicts (12). However, in more recent work presenting a series of short narratives about a hypothetical nonpolitical target, a retraction fully countered the negative effects from accusations of immoral behavior (13). A second set of studies using a similar approach likewise found no evidence for CIEs and saw some evidence of overcorrection, even under conditions where other presented information could support the truth of the retracted accusation (14). Our study will help to resolve these conflicting findings by testing whether CIEs limit effectiveness of corrections when political candidates are targeted by false misconduct accusations, across various candidate stimuli, on both single evaluations and binary choices, and in a context resembling real-world political information environments.

Another missing connection between CIEs and other approaches to researching misinformation is that many individual differences that predict vulnerability to misinformation and conspiracy theories in the real world have not been examined with respect to CIEs. Some work has found that CIEs in a causal inference paradigm are unchanged when excluding those who forget the retraction but are notably larger in people who report not believing the retraction, relative to those who claim to believe it (15). This suggests that individual differences in how people process and trust information could affect CIEs towards political candidates. Thus, predictors of real-world vulnerability to misinformation may also be relevant to understanding which people are more vulnerable to persistent effects of refuted misinformation when making sociopolitical decisions.

One factor that can impact a person's vulnerability to misinformation is their tendency to think analytically. Specifically, higher scores on the Cognitive Reflection Test (CRT), which measures the degree to which people use analytical reasoning to overcome intuitive but incorrect responses, predicts greater accuracy discernment, i.e. success in identifying false information (16). Digital literacy—factual knowledge about digital and legacy media—is another factor that predicts improved accuracy discernment (17, 18); this knowledge may enhance analytic thinking about online content and responsiveness to cues about the reliability of information. A self-report measure of epistemic beliefs also correlates with belief in conspiracy theories (19, 20) and with measures of headline accuracy discernment and willingness to share false content (21). Of the three subscales on this measure, having faith in one's intuitive feelings and believing that truth is political (i.e. defined by those in power) are associated with stronger endorsement of conspiracy theories, poorer accuracy discernment, and greater willingness to share false information. Requiring evidence as a basis for beliefs is associated with reduced willingness to share false content, less endorsement of conspiracy theories, and better accuracy discernment.

In contrast, other work has suggested that social and affective factors like partisanship and ideological biases play an important role in vulnerability to misinformation (22–24). Republicans and those scoring lower on actively open-minded thinking (AOT) show lower accuracy discernment, suggesting a tendency for ideological bias known as "myside bias." Partisanship and AOT are stronger predictors of accuracy discernment than analytic thinking measures (25), supporting an "integrative account" of misinformation vulnerability. Affective polarization—a strong preference for one's party over opposing partisans—is associated with a greater likelihood of sharing false content on Twitter, particularly among Republicans (26). Note that because our stimuli are largely devoid of partisan cues, our intention is not to test whether people are more influenced by false information that aligns with their own partisan identify (cf. 27). Our design instead allows us to test whether political orientation and/or polarization are associated with being more strongly influenced by compelling scandal narratives even when these are factually refuted.

Another potential vulnerability of practical interest is age. It is unclear whether older adults are more or less vulnerable to misinformation. Older adults tend to show declines in working memory (28) and associative memory (29), suggesting a possible impairment in processes associated with CIEs: integrating refutations into mental models and recalling that an accusation has been refuted. Consistent with this expectation, older adults consume and share higher levels of false content on real social media platforms (30, 31). Aging has also been associated in the lab with reduced explicit memory for trivia knowledge having been identified as false (6). However, in other laboratory research, older adults have shown either similar (32) or even *reduced* CIEs compared to young adults in the context of causal inference in a classic CIE paradigm (33), and performed *better* than young adults at explicitly discerning true from false news headlines (34, 35). These disparate findings could be explained by differences in the cognitive and/or socioemotional processes engaged in some laboratory tasks vs. real-world social media platforms, or by differences in the older adult populations that are being studied. Regardless, additional evidence on how age affects CIEs for accusations against hypothetical political candidates may help clarify these seemingly conflicting findings between real-world and laboratory settings.

Finally, we aim to achieve a preliminary understanding of differences in processing accusations versus factual refutations. Specifically, we examined whether recognition memory varies between accusations and refutations. Prior work suggests that negative information is motivationally salient in a variety of evaluative contexts, including in social impression formation (36), and that negative emotion enhances memory for associated stimuli (37). We hypothesized that the negative emotionality evoked by accusations causes accusation stimuli to be prioritized in memory and decision-making relative to matched neutral stimuli, while refutations will receive less or no prioritization. Beyond any direct effect of memory modulation on decision-making, prioritized processing of accusation stimuli is a possible mechanism by which refuted accusations may continue to bias decisions (cf. 38).

Results

We ran two large online behavioral experiments with American participants, aiming to recruit 500 participants per experiment.

The first study was exploratory, while the second was a preregistered replication in which candidate judgements were also measured 2 days later. Note that, to address points raised during peer review, some of the analyses highlighted here differ from the preregistered analysis plan; deviations from the preregistration are detailed in the Methods. In both experiments, participants first read introductory bios for all candidates (Fig. 1A) and then saw two mock social media posts for each candidate (Fig. 1B). Each set of social media posts was presented in one of three formats, which were varied within-participants: *corrected accusation* (post 1: Accusation; post 2: Refutation), *uncorrected accusation* (post 1: Accusation; post 2: Refutation Control), or *no accusation* (post 1: Accusation Control; post 2: Refutation Control). Accusation control and refutation control stimuli were topically and visually matched to their respective accusation and refutation stimuli but were designed to have little impact on candidate ratings. Participants saw an equal number of candidates in each of the three formats, and we counterbalanced across participants which candidates appeared in which format. Immediately after reading the posts, participants rated the candidate on a feeling thermometer (Fig. 1C). The analyses reported here focus largely on continued influence effects (CIEs), subtracting mean ratings for candidates with *no accusation* from mean ratings for candidates with *corrected accusations*. Note that more negative scores on CIEs indicate a larger decline in ratings for candidates exposed to accusations. Analyses of secondary outcomes, which look at the effect of accusations and corrections separately, are reported in Supplementary Information Results.

After reading posts about each candidate and providing a rating, participants then completed a series of self-report questionnaires (see Methods), as well as the MIST-20 headline accuracy judgment measure (39). These were followed by delayed ratings about each candidate, following the same procedure as the immediate ratings, as well as a delayed choice task (Fig. 1D), in which participants completed a series of binary choices about which of two candidates they would vote for in an election. In the choice task, we computed CIEs as the proportion of trials where a corrected accusation candidate was chosen over a no accusation candidate when candidates of those two types were being compared. Finally, memory was assessed via a recognition memory test (Fig. 1E). In experiment 1, all of these measures were collected in a single testing session, while in experiment 2, memory and a second round of delayed ratings and choices were collected 2 days later.

Behavioral main effects

Immediate and delayed CIEs

As expected based on pilot testing (see Supplementary Information Results), CIEs were present in the aggregate in experiment 1 on immediate ratings, $t(436) = -11.73$, $P < 0.0001$, $d = -0.56$ (Fig. 2A). CIEs persisted on ratings made after a short (~20–30 min) delay, $t(436) = -5.63$, $P < 0.0001$, $d = -0.27$ (Fig. 2B). CIEs were also evident on the choice task at a short delay in experiment 1, $t(436) = -5.48$, $P < 0.0001$, $d = -0.26$ (Fig. 2C). In preregistered analyses for experiment 2, CIEs were present on immediate ratings, $t(516) = -11.35$, $P < 0.0001$, $d = -0.50$ (Fig. 2D), on ratings made after a short delay, $t(516) = -6.17$, $P < 0.0001$, $d = 0.27$ (Fig. 2E), and on choices made after a short delay, $t(516) = -5.12$, $P < 0.0001$, $d = -0.23$ (Fig. 2F). In experiment 2, we also examined CIEs after a 2-day delay; these effects were present in preregistered analyses for both the delayed rating task, $t(401) = -5.45$, $P < 0.0001$, $d = -0.27$, and the delayed choice task, $t(401) = -4.32$, $P < 0.0001$, $d = -0.22$ (see Fig. S1). Contrary to our preregistered prediction, CIEs in experiment 2 did not differ between short-delay and long-delay measurements for the rating task, $t(401) = 1.17$, $d = 0.06$, or for the choice task, $|t|(401) < 1$, $d = -0.04$ (this issue is discussed further in Supplementary Information Results). Analyses of CIEs that show similar results for alternative exclusion criteria are reported in Table S1.

Comparison of initial and immediate ratings

To control for any effects of candidate demographic and biographical details, assignment of candidate to condition was counterbalanced between individuals. Mixed-effects models that account for initial ratings made prior to reading the social media posts, presented in Supplementary Information Results, confirm that CIEs were robust to differences in these initial ratings. Another approach to measuring effects of the mock social media posts is to compare immediate ratings after reading each social media post to the initial rating for that candidate. This approach also showed a clear decline from initial ratings to immediate post-story ratings for candidates with corrected accusations in experiment 1, $t(436) = -8.24$, $P < 0.0001$, $d = -0.39$, and in experiment 2, $t(516) = -8.19$, $P < 0.0001$, $d = -0.36$ (see Fig. S2).

CIEs and memory

We additionally find that CIEs do not appear to be a direct consequence of trials in which the accusation was remembered while the refutation was not, as would be implied by a simple version of the selective retrieval account of CIEs. That is, CIEs were still present when limiting the analysis to trials in which all four stimulus types (see Fig. 1B) were correctly categorized on the later recognition test, though many participants needed to be excluded from these analyses due to having no such trials in a given condition. On this recognition test, familiarity with a candidate was not sufficient to yield a correct response, as all stories involved candidates who participants had seen before; instead, participants needed to remember which two specific social media posts (accusation, accusation control, refutation, or refutation control) they had seen previously for a given candidate and which two they had not. When conditionalizing on memory, CIEs were present on immediate ratings in experiment 1, $t(319) = -5.94$, $P < 0.0001$, $d = -0.33$, and in experiment 2, $t(217) = -6.06$, $P < 0.0001$, $d = -0.41$. CIEs conditionalized on memory were also present on short-delay ratings in experiment 1, $t(319) = -5.79$, $P < 0.0001$, $d = -0.32$, on short-delay ratings in experiment 2, $t(217) = -4.48$, $P < 0.0001$, $d = -0.30$, and on long-delay ratings in experiment 2, $t(217) = -4.99$, $P < 0.0001$, $d = -0.34$.

Data from the choice task were less clear. Here, we limited the analysis to candidate pairings in which all four stimulus types were remembered correctly for both candidates, and to individuals for whom at least three choice trials relevant to CIEs met this criterion. This analysis shows a significant CIE at a short delay in experiment 1, $t(186) = -3.44$, $P = 0.0007$, $d = -0.25$, but not in experiment 2, whether at a short delay, $t(48) = -1.57$, $P = 0.124$, $d = -0.22$, or at a long delay, $|t|(48) < 1$, $d = -0.11$. Still, all effects were in the same direction as in the analysis not conditionalized on memory, with corrected accusation candidates less likely to be chosen than no accusation candidates, and with effect sizes that are largely comparable to the main analysis. To more systematically test whether the observed null effects may be due to a failure to remember the refutation, we compared whether the proportion of corrected accusation candidates chosen differs within the same individuals when the analyses are or are not

Fig. 1. Stimulus/task design for A) introductory bios, B) core stimuli, C) ratings (immediate and delayed), D) delayed choices, and E) memory test.

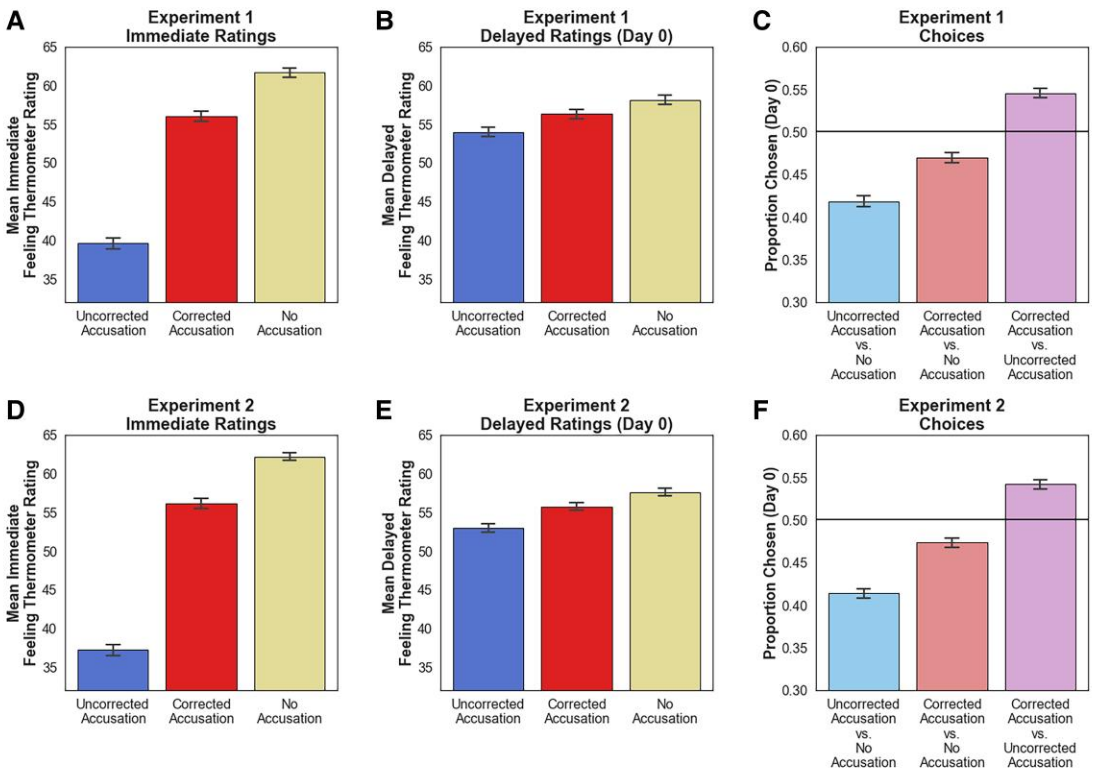

Fig. 2. Main effects of condition on A) experiment 1 immediate ratings, B) experiment 1 short-delay ratings, C) experiment 1 short-delay choices, D) experiment 2 immediate ratings, E) experiment 2 short-delay ratings, and F) experiment 2 short-delay choices. Effects shown in experiment 2 constitute a preregistered replication of effects observed in experiment 1. Error bars represent ±1 SE.

conditionalized on memory. No evidence for such a difference was found, whether in experiment 1, $t(186) = -1.49$, $d = -0.11$, $P = 0.14$, in experiment 2 at a short delay, $|t|(48) < 1$, $d = -0.06$, or in experiment 2 at a long delay, $|t|(48) < 1$, $d = -0.06$. Thus, the null effects for choices in experiment 2 are more likely due to a lack of power when conditionalizing on successful memory after a 2-day delay; we find no direct evidence supporting a role for memory failures in producing CIEs. Analyses of CIEs conditional on memory that show similar results for alternative exclusion criteria are reported in Table S2.

Individual difference predictors of CIEs

We examined bivariate correlations to determine which factors predict CIEs on the immediate feeling thermometer measure. We focus on immediate ratings because CIEs are larger and more reliable for this measure than for delayed ratings or choices. Specifically, as an estimate of reliability, we calculated CIEs for three subsets of the candidates, each composed of all candidates running for one of the three political offices, and examined whether CIEs were correlated across these three subsets. As described further in Supplementary Information Results and Table S3, these correlations were consistently and significantly greater than zero for immediate ratings in both experiments, but no significant positive correlations were observed for delayed ratings or choices. As might be expected based on this lack of internal consistency, no individual difference measures reliably predicted CIEs on delayed ratings and choices in either experiment (see Supplementary Information Results).

Three variables were significant predictors of immediate CIEs in both experiment 1 and experiment 2 (see Fig. 3; Table S4). Higher faith in intuition predicted stronger CIEs (Exp. 1: $r = -0.277$, $p_{FDR} < 0.0001$; Exp. 2: $r = -0.144$, $p_{FDR} = 0.013$), while higher digital literacy (Exp. 1: $r = 0.129$, $p_{FDR} = 0.018$; Exp. 2: $r = 0.111$, $p_{FDR} = 0.044$) and older age (Exp. 1: $r = 0.147$, $p_{FDR} = 0.008$; Exp. 2: $r = 0.123$, $p_{FDR} = 0.030$) predicted weaker CIEs. Correlation tables showing all bivariate correlations with immediate CIEs, as well as regressions examining effects of race/ethnicity, are presented in Table S4.

Regression analyses suggest that these three variables have largely independent effects on CIEs. Faith in intuition and age remain significant predictors of CIEs in both experiments in a multiple regression that includes all individual difference variables simultaneously and both are also selected as significant predictors in stepwise regressions, which we had originally preregistered as our individual differences analysis for experiment 2 (Tables S5 and S6). Digital literacy remains a predictor of CIEs in a multiple regression including all variables simultaneously, and is retained as a significant predictor in stepwise regressions, in experiment 2 but not experiment 1 (Tables S5 and S6).

We next examined whether any of the tested individual difference measures show reliably different correlation coefficients with CIEs compared to headline accuracy discernment. To address this question, we used tests of dependent correlation coefficients (40), accounting for the correlation between CIE and accuracy discernment measures. The most notable finding from these analyses (see Table S4) was that the effect of political party differed between measures (Fig. 4). Specifically, Republicans scored worse on headline accuracy discernment (Exp. 1: $r = -0.244$, $p_{FDR} < 0.0001$; Exp. 2: $r = -0.314$, $p_{FDR} < 0.0001$), but there was no effect of party on CIEs (Exp. 1: $r = -0.034$, $P = 0.48$; Exp. 2: $r = -0.013$, $P = 0.77$), and this difference was significant in both experiments (Exp. 1: $t = 3.60$, $p_{FDR} = 0.003$; Exp. 2: $t = 5.61$, $p_{FDR} < 0.0001$). The effects of political party on CIEs and MIST were also significantly different from each other in experiment 2, and

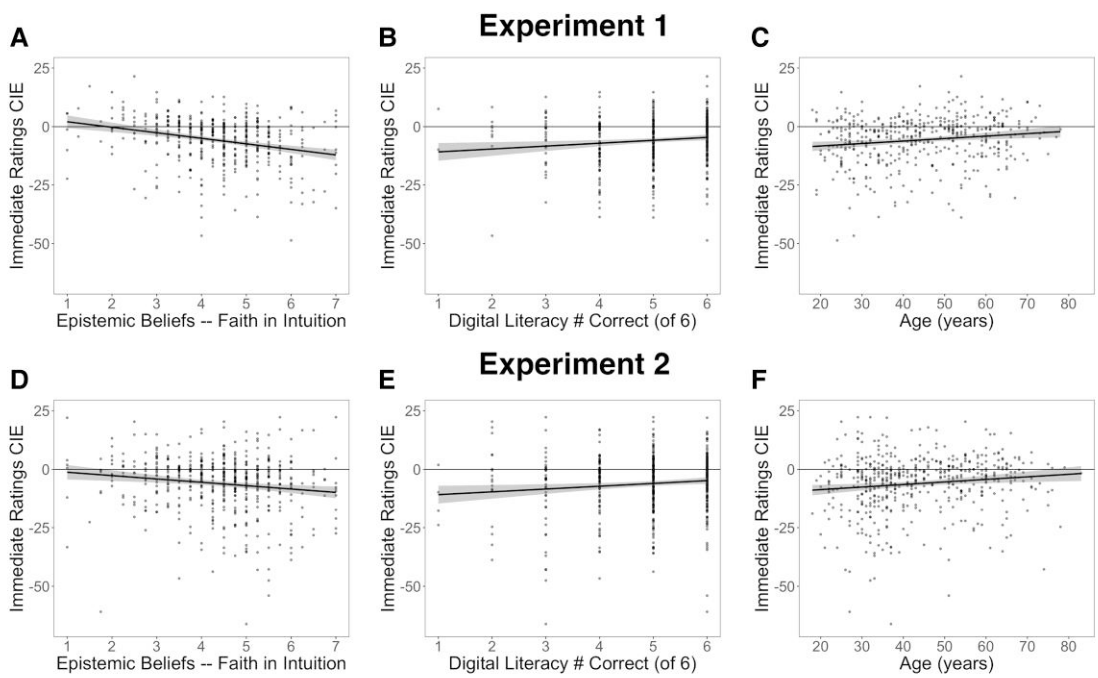

Fig. 3. Relationship between CIEs showing that A, D) higher self-reported faith in intuition is associated with larger CIEs, i.e. a larger drop in ratings in the corrected accusation condition relative to the no accusation condition, B, E) higher digital literacy is associated with reduced CIEs, and C, F) older adults show reduced CIEs, in A–C) experiment 1 and D–F) experiment 2. Shaded regions represent 95% CI.

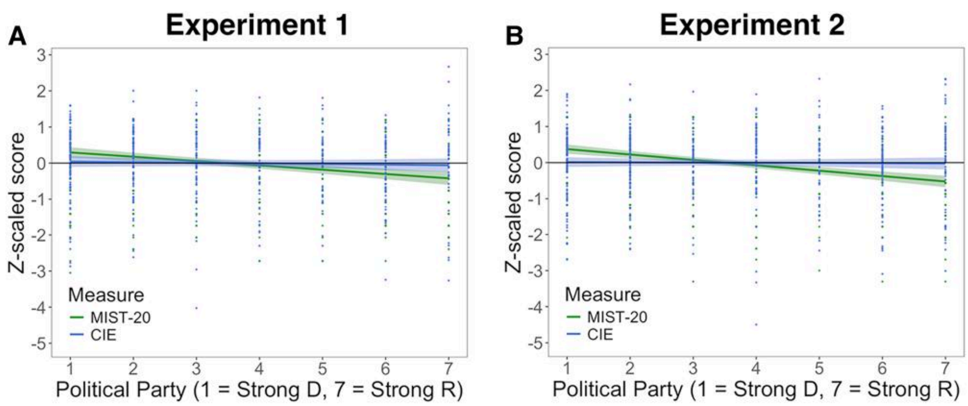

Fig. 4. Relationship between political party and both MIST-20 and CIE scores, showing that republicans perform more poorly on the MIST-20 than democrats, but do not show a larger CIE, in A) experiment 1 and B) experiment 2. Shaded regions represent 95% CI.

marginally different in experiment 1, when using an SUR regression approach (41) that we had originally preregistered as our analysis for this test for experiment 2 (Tables S5 and S6).

There were other differences between the correlations with CIEs and MIST, but these reflected a difference in degree of relationship rather than in whether a relationship is present at all. Specifically, digital literacy and all three epistemic belief subscales were more strongly correlated with the MIST than CIEs, but these variables showed at least marginal relationships with both CIEs and MIST. Individual differences analyses that show largely similar results (i.e. all significant effects reported above are at least marginal) for alternative exclusion criteria are reported in Tables S7–S12.

Memory

Finally, we examined whether explicit memory differs for accusations vs. refutations. Although CIEs are still present when participants remember all information, memory differences could nonetheless provide insight into how accusations and refutations are processed differently. We specifically examined whether the categorical benefit to memory for stimuli that should impact social impressions (i.e. accusations and refutations), relative to matched control stimuli lacking such impact, would differ for accusations vs. refutations. To do so, we ran a 2 (Impactfulness: Accusation/Refutation vs. Control) × 2 (Post: post 1 vs. post 2) repeated-measures ANOVA. In experiment 1, this analysis showed a main effect of impactfulness, $F(1, 403) = 23.89$, $P < 0.0001$, $\eta_p^2 = 0.056$, with impactful stimuli generally being remembered better than control stimuli. There was no main effect of post, $F(1, 403) < 1$, nor was there a reliable interaction between impactfulness and post, $F(1, 403) = 1.90$, $P = 0.17$, $\eta_p^2 = 0.005$, indicating that accusations benefit about as much as refutations in memory relative to matched control stimuli (Fig. 5A).

We examined this effect again in experiment 2, in which memory was tested 2 days after encoding rather than ∼30–40 min after

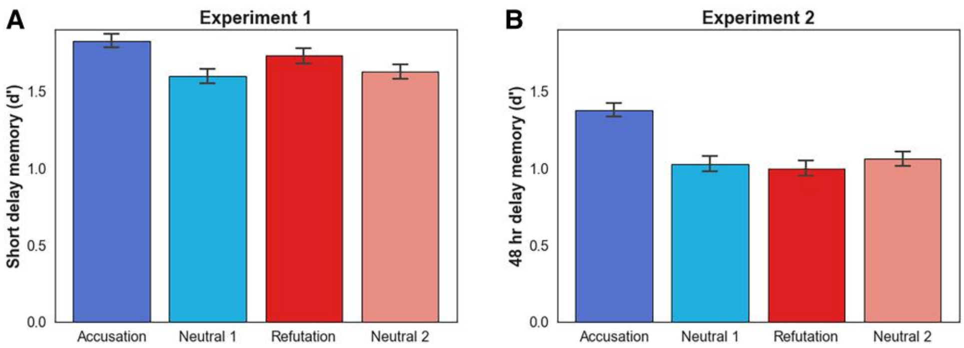

Fig. 5. Memory by stimulus condition. A) Experiment 1, memory after a short delay (~30 min). B) Experiment 2, preregistered replication showing memory after a longer (2-day) delay. Error bars represent ±1 SE.

encoding (Fig. 5B). Because emotional stimuli tend to affect memory consolidation more than immediate memory (42), we preregistered a prediction of a reliable interaction effect in experiment 2. Indeed, in experiment 2, we found a main effect of impactfulness, $F(1, 386) = 21.24$, $P < 0.0001$, $\eta_p^2 = 0.052$, a main effect of post, $F(1, 386) = 28.61$, $P < 0.0001$, $\eta_p^2 = 0.069$, and critically, an interaction between these variables, $F(1, 386) = 29.06$, $P < 0.0001$, $\eta_p^2 = 0.070$. The interaction indicates a strong memory benefit for accusation stimuli vs. accusation control stimuli, $t(386) = 7.22$, $P < 0.0001$, $d = 0.37$, but no advantage for refutation stimuli vs. refutation control stimuli, $|t|(404) < 1$, $d = -0.05$. Finally, we ran an additional 2 × 2 × 2 (impactfulness × post × experiment) mixed ANOVA to test whether the interaction between impactfulness and post differed as a function of the retention interval. This analysis showed a three-way interaction, $F(1, 789) = 10.17$, $P = 0.001$, $\eta_p^2 = 0.013$, indicating that the memory benefit for accusation stimuli was greater with a longer retention interval, as well as a main effect of experiment, $F(1, 789) = 141.74$, $P < 0.0001$, $\eta_p^2 = 0.15$, reflecting poorer memory at the longer retention interval, an interaction between post and experiment, $F(1, 789) = 11.53$, $P < 0.0007$, $\eta_p^2 = 0.014$, and other lower-order effects repeating those reported above.

Discussion

In this study, we report a novel approach studying how false accusations can harm politicians' reputations, even after those allegations have been factually refuted. We found that hypothetical candidates targeted by false but refuted accusations were rated more negatively than those never accused. The negative impression persisted regardless of whether people rated the candidates immediately, after a short delay, or even after 2 days, consistent with prior work (e.g. 33). We also discovered reliable individual differences in CIEs on immediate ratings. The tendency to maintain negative impressions is stronger for those who report relying more on intuition, people with lower digital literacy skills, and younger individuals. Interestingly, political affiliation did not affect this tendency. Finally, although we did not find direct evidence that CIEs result from selective retrieval of accusations, we did find that people have better memory for accusations than refutations after a 2-day delay. We speculate that emotionally charged information (like accusations) simultaneously impacts both memory consolidation processes and decisions about political candidates.

Our findings replicate and extend a prior study testing impressions of mock political candidates accused of misconduct (10), while differing from other recent studies examining the formation of social impressions about ordinary individuals (13, 14). It has been suggested (43), based on the finding in (10) that CIEs may be eliminated when decisions are both preceded by explicit deliberation and are about a same-party political candidate, that CIEs towards politicians may be limited to when there is a reluctance to update negative views about opposing-party politicians. In contrast, we conclude that in a nonpartisan political context, CIEs towards politicians are robust, and that CIEs apply more broadly in sociopolitical impression formation than other recent work has suggested (13, 14, 43). Our paradigm does differ in other ways, such as in having two stimuli about each of a large number of candidates rather than many stimuli about one targeted individual. Our design may be more ecologically valid, and under these conditions the cognitive bandwidth available for impression updating may be reduced, increasing the relative impact of emotionally salient accusations. In any case, we find strong evidence that people maintain negative opinions about politicians even when allegations are demonstrably false.

We find as well that the degree to which retracted information continues to influence social impressions varies reliably between individuals. Specifically, those who report relying more on intuition in choosing what to believe were more affected by debunked information. These results build on previous work suggesting that an increased reliance on emotion, whether pre-existing or induced by task context, can lead to increased belief in false information (44). In contrast, those with higher digital literacy—an acquired understanding of how social media platforms work on a mechanistic level—are less influenced by false accusations. Understanding how social media works might make people more skeptical of the content they encounter online, making them more analytical about what they read, and potentially more willing to trust refutations from a verified source (cf. 15). Another question for future work is to examine how cognitive abilities, like working memory, can affect people's ability to spot false information, particularly in political contexts (cf. 7–9). Ultimately, these findings are most notable in showing that CIEs vary systematically with measures (i.e. reliance on intuition and digital literacy) that also predict the ability to explicitly discern true from false headlines (e.g. 17–19, 44).

We find that partisan orientation is unrelated to the magnitude of CIEs. These results are interesting in the context of work emphasizing the role of partisan orientation and other socioemotional factors in headline accuracy discernment, in addition to analytic thinking, forming the "integrative account" of misinformation vulnerability (24, 25). We find that while Republicans were more likely to believe false headlines in the MIST, they were just as willing as Democrats to change their impressions of novel political candidates after an accusation was debunked. Our results suggest that Republicans are not inherently more vulnerable to making decisions based on debunked false information. On the MIST, true headlines reference specific recent issues and events, and false headlines are generated to resemble real-world misinformation. In recent years, American conservatives have been exposed to more false information than similarly situated liberals (30, 45, 46). In contrast, all participants in our study hear the exact same information about the fictional politicians. Thus, our data imply that the conservative worldview is not inherently more credulous or dogmatic (cf. 47).

We also observed that older adults consistently show reduced CIEs, replicating another recent study that used nonpolitical stimuli (33). Older adults have previously been shown to have worse explicit memory that a piece of information had been tagged as a myth (6); however, given that we find age predicts reduced CIEs on immediate judgments (where the stimuli are likely still in working memory), and that CIEs are present even when people remember all stimuli later, it seems reasonable to assume that the social impression judgments in our study rely to a substantial degree on processes other than explicit memory. Our results align with research on the age-related positivity bias (48), including recent work showing that older adults are more forgiving of selfish behavior than young adults (49). Older adults may be less inclined to judge a politician harshly based solely on an accusation, especially after evidence has refuted the allegation. Our result also aligns with prior work demonstrating better headline accuracy discernment in older adults (34, 35). However, it is possible that older adults who participate in online studies may differ from the general population, whether in cognitive ability (e.g. 50), digital savviness, or some other factor. Future work will need to address this limitation. It is also possible that at a delay longer than 2 days, memory differences with age would make older adults more vulnerable to CIEs than what we observed. Regardless, it remains unclear why older adults share more false information on social media than their younger counterparts (30, 31).

Finally, we found evidence that people process accusations preferentially relative to refutations. In our second experiment, participants remembered accusations better than refutations after a 2-day delay. These results differed from our first experiment, which used a shorter retention interval and showed no significant difference in memory between accusations and refutations. This result is consistent with accusations triggering greater emotional arousal, thereby enhancing memory consolidation (42). We also found tentative evidence, based on a correlation between CIEs and memory for accusations in experiment 2 (see Supplementary Information Results; Fig. S3), that would more directly support the interpretation that a common mechanism such as emotional arousal strengthens both CIEs and memory consolidation. Still, more research is needed to establish a role for preferential processing of emotionally evocative accusations in decision-making about political candidates.

One limitation of this study is that our newly developed digital literacy measure has not been rigorously validated from a psychometric standpoint. Nonetheless, our measure offers a more relevant assessment of digital literacy than other existing options. In one prior study (18), only one of four media literacy measures —specifically, information literacy—predicted how well people judged headline accuracy. This measure was notable for being the only one of the four measures to assess factual knowledge, but it does not focus on digital media. To our knowledge, the only measure relating factual digital literacy knowledge to misinformation vulnerability is a single question asking about the Facebook news feed (17). We used this question as a starting point but added additional questions, as we believe digital literacy is a broad enough construct to require multiple questions to assess robustly, following (51). As described further in Supplementary Information Results, we confirmed using Cronbach's alpha that our measure had acceptable reliability in a pilot sample, but more work is needed to further validate and refine this measure. Still, we expect that an ongoing challenge in measuring digital literacy is keeping factual knowledge questions accurate and relevant as the online media landscape rapidly evolves.

We expect this work to inspire new avenues for future research. Our work is a proof of concept that it is possible to measure how strongly refuted false information continues to affect decisions about political candidates at the individual level. Moreover, our method focuses on how misinformation influences more consequential political decisions than the judgment of whether a piece of information is accurate, as even information known to be inaccurate can still impact decisions. Finally, this approach allows for measurement of misinformation's impact on political decisions over time, not just immediately after exposure. Our method can be adapted to study misinformation in other areas, including false claims about the COVID-19 vaccine or medical treatments. This work is a critical step towards achieving a better understanding of how misinformation affects decisions and developing interventions that mitigate the real harms of false information by considering not just beliefs but also tangible behaviors.

Methods

Stimuli

We originally developed 36 fictional political candidates but reduced this number to 27 for the current set of studies to keep the experiment under 60 min. Both versions are available in our OSF repository (https://osf.io/gjpr9/). Each candidate's story was inspired by a real politician targeted by a false accusation that was later debunked by media sources, e.g. FactCheck.org. The stories covered nine scandal categories: bribery, electoral fraud/interference, embezzlement/self-dealing, racism, abuse of power or discretion, sexual harassment, foreign influence, financial fraud, and pedophilia/bestiality.

Our stimulus set included 18 men and 9 women. Of these candidates, 18 were white, and 9 were nonwhite (equally split between black, Hispanic, and Asian). This distribution roughly matches the actual demographics of US political officeholders while allowing each of the three counterbalancing groups to have an equal proportion of candidates across race and gender. Candidates were evenly split between running for US Senate, state governor, and US House positions. We tried to match the fictional candidates to their real-life inspirations in terms of race, gender, age, and political office. However, some deviations were necessary because white men were overrepresented in the real-life sample relative to our intended distribution.

We constructed a set of core stimuli for each candidate, consisting of an introductory biography and four social media posts: accusation, refutation, accusation-matched control, and refutation-matched control. The bios were loosely based on the real politician who inspired the story and typically included vague policy positions but avoided identifiable details. Each bio featured a campaign-style banner with the candidate's face, name, and prospective political office. The social media posts included this banner, along with a blurred stock photo image as a thumbnail with a blurred poster name next to it, a brief description of the story, and a headline and stock photo image previewing a hypothetical linked article on Facebook.

Each participant saw two social media posts about a given political candidate. The first post was either an accusation or accusation-matched neutral control post, while the second was either a refutation or a refutation-matched neutral control post. To minimize possible confounds, the meaningful posts (accusations or refutations) were matched to the paired control stimuli on general topic, article preview image (but with different text captions), and the blurred author thumbnail. For the second post (refutation or neutral), the thumbnail was a blurred logo and blurred name of a reputable news source intended to be politically neutral (e.g. Reuters) with an unblurred blue checkmark next to it. When refuting the accusation, these posts gave a clear causal account of the allegation by explaining how it came about via either a misunderstanding or deliberate fabrication.

CIEs were apparent in the first version of our stimulus set (see Supplementary Information Results). However, we pilot-tested a series of modifications to ensure that for each candidate, uncorrected accusations negatively impacted post-story ratings, refutations countered this negative impact relative to refutation controls, and neutral posts did not change opinions much in either direction. In this process, we did not modify refutations to make them *less* effective at correcting false allegations. See Supplementary Information Methods for additional details about stimulus design and pilot testing.

Experiment 1—Exploratory study

Participants

We planned to recruit 500 participants using the CloudResearch MTurk Toolkit. We chose the sample size based on our budget and preliminary power analysis, which showed that we could detect a bivariate correlation of at least $r > 0.125$ with 80% power. Recruitment was stratified by age, with equal samples targeted from five age brackets (18–29, 30–39, 40–49, 50–59, 60+). Participants were paid $8.00 for completing the study.

We excluded participants if they: failed simple attention checks at the start of this study, were removed from the CloudResearch Approved Participants list by July 2024, reported an age that did not match their birth year on file with CloudResearch (i.e. off by more than 1 year), or had a median time of less than 1.5 s reading either the first or second post, which would imply an implausibly fast reading speed of about 1,800 words per minute. A total of 499 participants completed experiment 1. In total, we excluded 62 participants from all analyses, some of whom failed on multiple criteria: 3 failed simple attention checks, 15 were removed from the CloudResearch Approved list, 35 reported their age inconsistently, 11 had below-threshold viewing time on post 1, and 18 had below-threshold viewing time on post 2. After exclusions, we had a final sample of 437 participants.

We also report data from our primary analyses with minimal exclusions in Supplementary Information Results. Based on peer review feedback, we modified the initial exclusion criteria noted in our preregistration. Data using our original preregistered exclusion criteria (below-chance performance on either the MIST or digital literacy measures) are also reported in Supplementary Information Results. Note that in experiment 2, only those who met our original exclusion criteria were invited back for the delayed judgments and memory test. To enable comparisons in memory performance across the two experiments, we excluded from analyses of memory performance in both experiments individuals who failed our original exclusion criteria. This yielded an additional 24 exclusions in experiment 1: 22 individuals who were below-chance on the MIST, 1 who was below-chance on the digital literacy measure, and 1 who failed on both measures. Results were similar without making these additional exclusions. Finally, nine additional individuals were excluded from analyses of memory performance because they scored below chance on the memory test.

Procedure

Study procedures were reviewed and approved by the University of Pennsylvania IRB #8 (protocol 844066) as consistent with the Declaration of Helsinki and the US federal Common Rule (45 CFR part 46) regulating human subjects research. Informed consent was obtained from participants at the beginning of each study session. Participants first saw introductory bios (Fig. 1A) for all 27 candidates in random order. Immediately after each bio, they rated how much they liked each candidate on a 0–100 feeling thermometer scale (initial ratings). Participants then saw two mock social media posts about each candidate (Fig. 1B). Candidates were evenly divided between the three conditions (*corrected accusation*, *uncorrected accusation*, or *no accusation*), with candidate assignment to these conditions counterbalanced across participants. This is the only stimulus feature that was varied systematically and counterbalanced. After seeing each pair of posts, participants again rated each candidate on the 0–100 feeling thermometer scale (immediate post-story ratings; Fig. 1C).

After viewing posts about all 27 candidates, participants completed several questionnaires: the Cognitive Reflection Test (CRT-2) (52); epistemic beliefs measure (19); two affective polarization measures—a dictator game, following (53), and a partisan feeling thermometer, following (54); a belief superiority measure (55); and the MIST-20 headline accuracy discernment scale (39). We also included a novel digital literacy measure, building on (17, 18); the digital literacy measure included nine multiple-choice questions with four response options each to examine factual knowledge of user experience and content moderation on social media platforms. More specifically, the digital literacy measure includes three questions about specific platforms (Facebook, Twitter, and TikTok) and three questions about specific concepts (phishing, blocking, and tagging). We computed digital literacy scores as the percent correct across all six items. See additional details in Supplementary Information Results and Appendix. All measures that were collected are listed here.

Following these questionnaires, participants made two judgments about each candidate. First was a choice task. Here, each candidate was paired with each of the other candidates running for the same political office (Governor, US Senate, or US House), and participants chose on each trial which candidate they preferred by clicking on the banner for that candidate (Fig. 1D). Participants made 36 choices per office, totaling 108 choice trials. Second, participants completed a delayed feeling thermometer rating for each candidate, in random order, again prompted only by the candidate banner. Third, in a recognition memory test (Fig. 1E) they saw all four stimuli (accusation, accusation control, refutation, refutation control) for each of the 27 candidates, and provided ratings using a 5-point scale ranging from "Definitely new" to "Definitely old." Finally, participants answered demographic questions about their age, race/ethnicity, political affiliation, and ideology (both following 7-point ANES survey format), education, income, household size, city, state, and ZIP code, and an open-ended feedback question.

Data analysis

On feeling thermometer rating measures, we computed a CIE score for each individual by subtracting the individual's mean score for *no accusation* candidates from the mean score for *corrected* *accusation* candidates. Note that this means that negative scores indicate the presence of CIEs, with larger CIEs yielding more negative scores. This approach was chosen so that negative values indicate greater influence of false information for both CIEs and headline accuracy discernment. For choices, CIEs were computed as the proportion of trials for which a *corrected accusation* candidate was chosen, from trials in which the choice was between a *corrected accusation* candidate and a *no accusation* candidate. Here again, lower scores indicate greater CIEs.

We also computed an affective polarization measure for those who expressed a preference between the Republican and Democratic parties; those with a weak preference were included, but this measure could not be computed for those who expressed no partisan preference. We computed the difference in the amount shared (out of $10) with in-party vs. out-party targets in the dictator game, as well as the difference between in-party vs. out-party feeling thermometer ratings. The Z-score of the in-party vs. out-party difference for each measure was computed across the sample, and the final measure of affective polarization was the participant's average Z-score across the two measures. Based on the finding in experiment 1 that greater affective polarization predicted larger CIEs, our preregistration included a mixed-effects analysis examining whether such an effect interacted with perceived candidate ideology and participants' political ideology on the item level. The overall effect of affective polarization did not replicate in experiment 2, however, so this analysis is reported only in Supplementary Information Results.

For demographic variables of education and income, we converted categorical responses to ordinal numbers for inclusion in regressions, as described in our preregistration. Education was converted to years of education such that "Doctoral degree" = 20, "Master's degree" = 18, "Bachelor's degree" = 16, "Associate's Degree" = 14, "Some college" = 13, "High school diploma" = 12, and "Have not finished high school" = 11. A small number of respondents chose "Other," and these responses were coded on an ad hoc basis, with vocational/technical school graduates and those currently in college coded as 13, and a respondent indicating a professional degree coded as 19. Income was converted to a category, consistent with our preregistered analysis plan, with "Under $20,000" = 1, "$20,000–$40,000" = 2, "$40,000–$75,000" = 3, "$75,000–$100,000" = 4, "$100,000–$500,000" = 5, and "Over $500,000" = 6.

We examined the correlations between CIE and MIST and the following measures: CRT % correct, Digital literacy score, Epistemic beliefs (Faith in Intuition), Epistemic beliefs (Faith in Evidence), Epistemic beliefs (Truth is political), Political party (1–7 scale), Affective polarization, Belief superiority, Age, Gender (Male = 0, Female = 1), Education, and Income. Each set of correlation coefficients was corrected across these 12 measures using a false discovery rate (FDR) correction (56), implemented with the R *p.adjust* function. Effects of race and ethnicity were examined (using dummy codes for Black, Hispanic, and Asian identity) in separate multivariate regressions reported in Supplementary Information Results. To test whether specific predictor variables had a different relationship with CIEs vs. MIST, we used the Steiger's Z test for dependent correlation coefficients (40), accounting for the correlation between CIE and MIST measures, and applying FDR correction across all 12 tests.

In addition to these bivariate correlations, we also examined the significant predictors of CIEs and MIST in multiple linear regressions that included all predictor variables, as well as in regression models selected through a stepwise procedure via the stepAIC algorithm in R. Note that the examination of the bivariate correlations and full multiple regression model was a deviation from our preregistered analysis plan, which only specified stepwise regression (see Supplementary Information Results for further explanation of this change made in response to peer review feedback).

In the memory test, for any given candidate and participant, two stimuli were actually "old" and two were actually "new," with the specific assignment of candidate to condition varying based on counterbalancing. The memory test was structured such that only one of the accusation or the accusation control stimuli, and only one of the refutation or refutation control stimuli, would appear in the first half of the memory test, with other stimuli presented in the second half of the memory test. Specific stimuli presented in each half of the test were counterbalanced. For analyses of memory data by condition, we only used data from the first half of the test, to avoid contamination of memory estimates when participants had already seen a matched stimulus earlier in the memory test. Both "Definitely old" and "Probably old" were counted as "old" responses, while "Definitely new" and "Probably new" were counted as "new" responses, and "Not sure" responses were excluded from analysis. Hit rates and false alarm rates were computed for each of the four types of stimuli, with up to 54 trials per condition, and d' scores for each stimulus type were calculated with the log-linear correction applied (57). In the separate set of analyses in which candidate impressions were computed based only on trials for which all stimuli were remembered accurately, data from the full memory test data were used, with four stimuli per candidate.

Experiment 2—Replication study

Participants

We planned to recruit 500 participants again using CloudResearch MTurk Toolkit, stratifying by age across five brackets (18–29, 30–39, 40–49, 50–59, 60+). Due to a data collection error, we recruited an additional set of ∼100 participants in the 30–39 age bracket and a smaller than anticipated sample in the 18–29 age bracket. Participants received $7.50 for completing the first part of the study and an additional $5.00 if they returned for the second part of the study 2 days later.

Of the 561 participants, we excluded data from 44, yielding a sample of 517 participants. Our exclusions mirrored those from our first study, with some participants again failing multiple criteria: 1 failed simple attention checks, 11 were removed from the CloudResearch Approved list, 17 reported their age inconsistently, 12 had below-threshold viewing time on post 1, and 12 had below-threshold viewing time on post 2. As noted above, these exclusion criteria represent a change from our preregistered plan. However, for the memory test and other long-delay measures, we did not invite back 46 participants who did not meet our original exclusion criteria. This included 42 individuals who scored below chance on the MIST, 3 who scored below chance on the digital literacy measure, and 1 who scored below chance on both. After these exclusions and natural attrition, we had 402 individuals who completed both study sessions. For memory performance measures, we excluded another 15 participants who scored below chance on the memory test.

Procedure

Data collection and analysis for experiment 2 were similar to experiment 1, with the preregistered plan for experiment 2 available at https://osf.io/ubw6m. We used the same measures as in experiment 1, except we removed city and state demographic questions that duplicated ZIP code. We collected three exploratory individual difference measures: the Questionnaire of Cognitive and Affective Empathy (58), which includes separate cognitive empathy and affective empathy components, the 12-item abbreviated Intolerance of Uncertainty (IUS) scale (59), and the 10-item AOT scale (25, 60). Multiple comparison correction for correlation coefficients with the 12 measures repeated between the two studies was applied only across those 12 measures. In analyses of the additional exploratory measures in experiment 2, we applied multiple comparison correction across all 16 measures. Delayed choice and delayed rating measures were collected both at the end of the first experimental session and the beginning of the second experimental session. The recognition memory test was shifted to the end of the second session and was not administered in the first session.

Acknowledgments

We acknowledge Thomas Hogeboom and Aysha Gsibat for assistance in stimulus development and pilot testing.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This work was funded by two grants (in 2020 and 2021) from the Facebook/Meta Research Foundational Integrity research program. Facebook/Meta had no control over the research design, analysis, or publication decisions related to this work.

Preprints

Preliminary versions of this article were posted on the PsyArXiv preprint server at https://osf.io/preprints/psyarxiv/5dmt4.

Data Availability

Stimulus materials, anonymized raw data, processed data, and analysis code are available via our Open Science Foundation (OSF) repository (61).

References

1 Brady WJ, Wills JA, Jost JT, Tucker JA, Van Bavel JJ. 2017. Emotion shapes the diffusion of moralized content in social networks. *Proc Natl Acad Sci U S A*. 114:7313–7318.

2 Vosoughi S, Roy D, Aral S. 2018. The spread of true and false news online. *Science*. 359:1146–1151.

3 Johnson HM, Seifert CM. 1994. Sources of the continued influence effect: when misinformation in memory affects later inferences. *J Exp Psychol Learn Mem Cogn*. 20:1420–1436.

4 Chan MS, Jones CR, Hall Jamieson K, Albarracín D. 2017. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci*. 28:1531–1546.

5 Ecker UKH, Lewandowsky S, Tang DTW. 2010. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Mem Cognit*. 38:1087–1100.

6 Swire B, Ecker UKH, Lewandowsky S. 2017. The role of familiarity in correcting inaccurate information. *J Exp Psychol Learn Mem Cogn*. 43:1948–1961.

7 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.

8 Brydges CR, Gignac GE, Ecker UKH. 2018. Working memory capacity, short-term memory capacity, and the continued influence effect: a latent-variable analysis. *Intelligence*. 69:117–122.

9 Sanderson JA, Gignac GE, Ecker UKH. 2021. Working memory capacity, removal efficiency and event specific memory as predictors of misinformation reliance. *J Cogn Psychol*. 33:518–532.

10 Thorson E. 2016. Belief echoes: the persistent effects of corrected misinformation. *Polit Commun*. 33:460–480.

11 Ross L, Lepper MR, Hubbard M. 1975. Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *J Pers Soc Psychol*. 32(5):880–892.

12 Steblay N, Hosch HM, Culhane SE, McWethy A. 2006. The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: a meta-analysis. *Law Hum Behav*. 30:469–492.

13 Ecker UKH, Rodricks AE. 2020. Do false allegations persist? Retracted misinformation does not continue to influence explicit person impressions. *J Appl Res Mem Cogn*. 9:587–601.

14 Mickelberg A, *et al*. 2024. Does mud really stick? No evidence for continued influence of misinformation on newly formed person impressions. *Collabra Psychol*. 10:92332.

15 O'Rear AE, Radvansky GA. 2020. Failure to accept retractions: a contribution to the continued influence effect. *Mem Cognit*. 48: 127–144.

16 Pennycook G, Rand DG. 2019. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*. 188:39–50.

17 Sirlin N, Epstein Z, Arechar AA, Rand DG. 2021. Digital literacy is associated with more discerning accuracy judgments but not sharing intentions. *Harv Kennedy Sch Misinformation Rev* 2. https://doi.org/10.37016/mr-2020-83.

18 Jones-Jang SM, Mortensen T, Liu J. 2021. Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *Am Behav Sci*. 65:371–388.

19 Garrett RK, Weeks BE. 2017. Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. *PLoS One*. 12:e0184733.

20 Young DG, Maloney EK, Bleakley A, Langbaum JB. 2022. I feel it in my gut:" epistemic motivations, political beliefs, and misperceptions of COVID-19 and the 2020 U.S. Presidential election. *J Soc Polit Psychol*. 10:643–656.

21 Rudloff JP, Appel M. 2023. When truthiness trumps truth: epistemic beliefs predict the accurate discernment of fake news. *J Appl Res Mem Cogn*. 12:344–351.

22 McLoughlin KL, Brady WJ. 2023. Human-algorithm interactions help explain the spread of misinformation. *Curr Opin Psychol*. 56:101770.

23 Van Bavel JJ, Rathje S, Vlasceanu M, Pretus C. 2023. Updating the identity-based model of belief: from false belief to the spread of misinformation. *Curr Opin Psychol*. 56:101787

24 Van Bavel JJ, *et al*. 2021. Political psychology in the digital (mis)information age: a model of news belief and sharing. *Soc Issues Policy Rev*. 15:84–113.

25 Roozenbeek J, *et al*. 2022. Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgm Decis Mak*. 17:547–573.

26 Osmundsen M, Bor A, Vahlstrup PB, Bechmann A, Petersen MB. 2021. Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *Am Polit Sci Rev*. 115:999–1015.

27 Pereira A, Harris E, Van Bavel JJ. 2023. Identity concerns drive belief: the impact of partisan identity on the belief and dissemination of true and false news. *Group Process Intergroup Relat*. 26: 24–47.
28 Bopp KL, Verhaeghen P. 2005. Aging and verbal memory span: a meta-analysis. *J Gerontol B Psychol Sci Soc Sci*. 60:P223–P233.
29 Old SR, Naveh-Benjamin M. 2008. Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychol Aging*. 23(1):104–118.
30 Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazar D. 2019. Fake news on twitter during the 2016 U.S. Presidential election. *Science*. 363:374–378.
31 Guess A, Nagler J, Tucker J. 2019. Less than you think: prevalence and predictors of fake news dissemination on Facebook. *Sci Adv*. 5:eaau4586.
32 Guillory JJ, Geraci L. 2010. The persistence of inferences in memory for younger and older adults: remembering facts and believing inferences. *Psychon Bull Rev*. 17(1):73–81.
33 Miller AL, Wissman KT, Peterson DJ. 2022. The continued influence effect: examining how age, retraction, and delay impact inferential reasoning. *Appl Cogn Psychol*. 36:708–723.
34 Brashier NM, Schacter DL. 2020. Aging in an era of fake news. *Curr Dir Psychol Sci*. 29:316–323.
35 Arechar AA, *et al*. 2023. Understanding and reducing online misinformation across 16 countries on six continents. *Nat Hum Behav*. 7:1502–1513
36 Rozin P, Royzman EB. 2001. Negativity bias, negativity dominance, and contagion. *Pers Soc Psychol Rev*. 5:296–320.
37 Kensinger E, Corkin S. 2004. Two routes to emotional memory: distinct neural processes for valence and arousal. *Proc Natl Acad Sci U S A*. 101:3310–3315.
38 Zajonc RB. 2000. Feeling and thinking: closing the debate over the independence of affect. In: Forgas JP, editor. *Feeling and thinking: the role of affect in social cognition*. Cambridge UK; New York: Cambridge University Press. p. 31–58.
39 Maertens R, *et al*. 2023. The Misinformation Susceptibility Test (MIST): a psychometrically validated measure of news veracity discernment. *Behav Res Methods*. 56:1863–1899.
40 Steiger JH. 1980. Tests for comparing elements of a correlation matrix. *Psychol Bull*. 87:245–251.
41 Zellner A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J Am Stat Assoc*. 57:348–368.
42 McGaugh J. 2004. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu Rev Neurosci*. 27:1–28.
43 Ecker UKH, *et al*. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nat Rev Psychol*. 1: 13–29.
44 Martel C, Pennycook G, Rand DG. 2020. Reliance on emotion promotes belief in fake news. *Cogn Res Princ Implic*. 5:47.
45 Mosleh M, Rand DG. 2022. Measuring exposure to misinformation from political elites on twitter. *Nat Commun*. 13:7144.
46 Lasser J, *et al*. 2022. Social media sharing of low-quality news sources by political elites. *PNAS Nexus*. 1:pgac186.
47 Van der Linden S, Panagopoulos C, Azevedo F, Jost JT. 2021. The paranoid style in American politics revisited: an ideological asymmetry in conspiratorial thinking. *Polit Psychol*. 42:23–51.
48 Carstensen LL. 2006. The influence of a sense of time on human development. *Science*. 312:1913–1915.
49 Lempert KM, *et al*. 2022. Aging is associated with maladaptive episodic memory-guided social decision-making. *Proc Natl Acad Sci U S A*. 119:e2208681119.
50 Ogletree AM, Katz B. 2021. How do older adults recruited using MTurk differ from those in a national probability sample? *Int J Aging Hum Dev*. 93:700–721.
51 Diamantopoulos A, Sarstedt M, Fuchs C, Wilczynski P, Kaiser S. 2012. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *J Acad Mark Sci*. 40:434–449.
52 Thomson KS, Oppenheimer DM. 2016. Investigating an alternate form of the cognitive reflection test. *Judgm Decis Mak*. 11:99–113.
53 Iyengar S, Westwood SJ. 2015. Fear and loathing across party lines: new evidence on group polarization. *Am J Pol Sci*. 59: 690–707.
54 Druckman JN, Levendusky MS. 2019. What do we measure when we measure affective polarization? *Public Opin Q*. 83:114–122.
55 Toner K, Leary MR, Asher MW, Jongman-Sereno KP. 2013. Feeling superior is a bipartisan issue: extremity (not direction) of political views predicts perceived belief superiority. *Psychol Sci*. 24: 2454–2462.
56 Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 57:289–300.
57 Hautus MJ. 1995. Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behav Res Methods Instrum Comput*. 27:46–51.
58 Reiners RLEP, Corcoran R, Drake R, Shryane NM, Völlm BA. 2011. The QCAE: a questionnaire of cognitive and affective empathy. *J Pers Assess*. 93:84–95.
59 Carleton RN, Norton PA, Asmundson GJG. 2007. Fearing the unknown: a short version of the intolerance of uncertainty scale. *J Anxiety Disord*. 21:105–117.
60 Baron J, Isler O, Yilmaz O. 2022. Actively open-minded thinking and the political effects of its absence. PsyArXiv. https://psyarxiv.com/g5jhp/, preprint: not peer reviewed.
61 Cohen MS, Halewicz V, Yildirim E, Kable JW. 2024. Continued influence of false accusations in forming impressions of political candidates. *OSF repository*, https://doi.org/10.17605/OSF.IO/GJPR9