The Role of Worldview Congruence in Misinformation Correction: A Bayesian Approach to Belief Updating

Greta Arancia Sanna (greta.sanna.23@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, WC1H 0AP, London, UK

Toby D. Pilditch(t.pilditch@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, WC1H 0AP, London, UK

David A. Lagnado (d.lagnado@ucl.ac.uk)
Department of Experimental Psychology, University College London,
26 Bedford Way, WC1H 0AP, London, UK

Abstract

Misinformation poses a growing challenge in society, particularly as people seem reluctant to revise discredited information. However, emerging research suggests that people's persistence in believing discredited information is sometimes rational, with individuals applying their own assumptions in a consistent way. In this experimental study, we employ a political vignette with a false accusation to show that even in politically charged contexts, people can correct misinformation and return their beliefs to baseline. Our results demonstrate that participants' belief updating generally aligns with Bayesian predictions, although with a more conservative approach. With regards to source evaluation, participants were more likely to downgrade the reliability of an accuser's claim when it conflicted with their political views but returned to their initial assessments after the correction. This suggests that while worldview can affect source evaluation at the individual level, this effect does not necessarily translate into a broader erosion of institutional credibility. This study enhances our understanding of how worldview influences belief revision and source evaluation, especially in politically sensitive contexts.

Keywords: Misinformation correction; motivated reasoning; source reliability; Bayesian updating.

Introduction

Misinformation is of increasing concern in many areas of social life. In the political domain, false and misleading claims can have lasting negative consequences on how people revise their beliefs and, ultimately, on how well democracy functions (Bastos & Mercea, 2019; Oyserman & Dawson, 2020; Ross & Rivers, 2018). While the effect may be overstated (Atlay et al., 2023), misinformation remains widespread, and effective interventions are urgently needed. This raises important questions about how to maximise the effectiveness of corrective messages, particularly given the persistent finding that people often continue to believe in misinformation even after it has been corrected (Anderson et al., 1980; Guillory & Geraci, 2010; Johnson & Seifert, 1994; Ross et al., 1975; Wilkes & Leatherbarrow, 1988; Wilkes & Reynolds, 1999).

This phenomenon, known as the Continued Influence Effect (CIE), refers to the tendency for misinformation to persist in people's beliefs despite clear and credible corrections (Anderson et al., 1980; Guillory & Geraci, 2010). Key theories in this domain suggest that individuals often struggle to update their beliefs, particularly when the correction does not offer a coherent alternative explanation that fits into their existing mental model (Connor Desai & Reimers, 2019; Johnson & Seifert, 1994; Rich & Zaragoza, 2016). In contrast, recent analyses of the CIE propose that, in certain cases, the effect may be rational, and failing to revise one's beliefs in light of new evidence is not always illogical (e.g., Connor Desai et al., 2020; Haselton et al., 2009; Gershman, 2019; Pilgrim et al., 2024). In line with these approaches, Sanna and Lagnado (2024) show that people are more likely to discount an initial claim when it is corrected by a source perceived as more reliable than the original claimant. Conversely, when the claim is corrected by a source perceived as less reliable than the original claimant, participants exhibit the CIE.

Connor Desai et al. (2020) argue that characteristics of the source of information are essential to the evaluation of content: if a source is likely to lie or make errors, it is reasonable to discount their claims. Therefore, in cases where the source of misinformation is perceived as more credible than the source of the correction it might be rational not to update one's beliefs (Jern, Chang, & Kemp, 2014). This fits with O'Rear and Radvansky's (2020) findings, where the CIE only arises if people do not believe in the correction. Therefore, in cases where people fail to revise their beliefs despite a correction, this might be due to the rational evaluation of source reliability rather than cognitive bias.

To test this line of reasoning, researchers have used Bayesian models to explore whether individuals 'correctly' adjust their beliefs based on new evidence and prior knowledge. Empirical studies have shown that people are indeed able to incorporate their assessment of a source's credibility when evaluating testimony (Harris et al., 2016; Harris & Hahn, 2009; Madsen, 2016; Merdes et al., 2021) and

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

sensibly alter their belief in a source's reliability if new information, whether contradictory or corroborating, is made available (Madsen et al., 2020). Connor Desai et al. (2020) apply a Bayesian network formalism, conceptualising the CIE through scenarios with contradictory testimonies, and varying the perceived reliability of sources. They demonstrate that on average people update their beliefs in line with a rational Bayesian model.

These studies highlight people's potential for rational belief updating, emphasising the importance of incorporating source reliability. However, one limitation is that they involve experimental settings that avoid emotionally charged scenarios or topics. This raises concerns about how well these results can be generalised to contexts that might provoke motivated reasoning, such as politically-charged environments.

To address this lacuna, recent research investigates how motivated reasoning affects misinformation correction. Motivated reasoning refers to the tendency for individuals to process new information in a way that supports their pre-existing beliefs and attitudes (Kunda, 1990). This is particularly relevant in political contexts, where people's pre-existing attitudes may lead them to resist corrections that contradict their worldview (Taber & Lodge, 2006). For example, Lewandowsky et al. (2005) demonstrated that participants who supported the 2003 Iraq invasion were more likely to persist in believing retracted pro-invasion news items. Ecker and Ang (2019) also found that political worldview influenced both misinformation acceptance and the effectiveness of corrections. Specifically, corrections were less effective when the misinformation was worldview-congruent, with the effect being particularly pronounced for conservatives.

While many studies, such as Nyhan and Reifler (2010) and Thaler (2024), have found evidence of motivated reasoning in response to misinformation and corrections, other research challenges this view. For instance, Pennycook and Rand (2019) argue that analytic reasoning plays a crucial role in how people process suspicious content, suggesting that they are capable of discounting corrected information even when it contradicts their worldview. Ecker et al. (2020) also failed to replicate accepted findings in this literature, observing no impact of political worldview on correction effectiveness. Given the mixed results regarding the impact of worldview consistency on correction effects, our study aims to further investigate the role of worldview in belief updating by using Bayesian network models.

In this paper, we explore the extent to which political worldview affects people's belief revision when information previously presented is corrected. We use Bayesian Networks as a normative benchmark for individuals' belief updating. We address two key research questions: (1) To what extent do individuals exhibit rational belief updating when confronted with conflicting evidence? and (2) How does political worldview influence belief updating in such contexts?

To answer these questions, we use a vignette in which a politician running for re-election is accused of bribery by a prosecutor, followed by a correction from a different prosecutor denying the claim. Alleged scandals are common topics of debate with public figures often struggling to discredit false claims. This makes it a useful domain to explore how people update their beliefs when confronted with contradictory claims of a political nature.

Bayesian Network Model

To compare people's belief updating against a normative standard we use Bayesian Networks (Pearl, 1988), which provide a formal approach to modelling the probabilistic relations between multiple variables. Bayesian networks (BNs) are made up of two components: (1) a directed acyclic graph that represents the probabilistic relations between variables, and (2) probability tables for each variable that capture the strength of these relations and allow for quantitative updating via Bayes rule. To incorporate the role of source reliability we adapt the Bayesian network introduced by Bovens and Hartmann (2003) (see also Fenton et al., 2013, Lagnado, 2021, for use of BNs to model witness reliability in the legal domain).

The Bayesian network for the vignette used in our study is shown in Figure 1. We use binary variables to represent the key aspects of the vignette: whether or not the politician took the bribe (Bribe); the claim made by the accuser (Claim1); the claim made by the corrector (Claim2); the reliability of the accuser (Rel1); and the reliability of the corrector (Rel2). Critically, each claim is a function of the status of the bribe (true or false), and of the reliability of the source making the claims (reliable or unreliable). The priors for Bribe, Rel1 and Rel2, and the conditional probability tables for Claim1 and Claim 2 will be derived from participant judgments. By using this BN model and parameterising it with participants' stated priors and conditional probabilities, we can compare participant's actual belief updating (given evidence of the claims) against the predictions of the BN model (see Methods section for more details).

Hypotheses

In our study we examined the effect of political motivation on belief updating by varying the affiliation of the politician in the vignette (Democrat, Republican, or Unspecified). By recruiting both Democrat and Republican participants we assessed whether worldview congruence affects belief updating. Based on previous research (Connor Desai et al., 2020; Sanna & Lagnado, 2024) we predicted:

H1: People would follow a rational pattern of belief updating when political motivation is irrelevant (e.g., in the neutral/unspecified condition of our study).

H2: In the politically charged conditions (e.g., where the affiliation of the politician accused of taking a bribe is stated) updating would be affected by worldview congruence.

In line with these hypotheses, we expected worldview congruence to influence belief revision. Specifically, we anticipated that in the neutral condition, participants would

update their beliefs rationally, correcting the corruption claim without bias. However, in politically charged contexts, we expected participants to align their belief updates more closely with their pre-existing political views. Thus, corrections were predicted to be more effective when the participant's political affiliation aligned with that of the politician in the vignette and less effective when they are in opposition.

H3: Finally, we predicted that the Bayesian network model would effectively capture both participants' tendency for internal consistency in belief updating and the influence of political worldview on these processes.

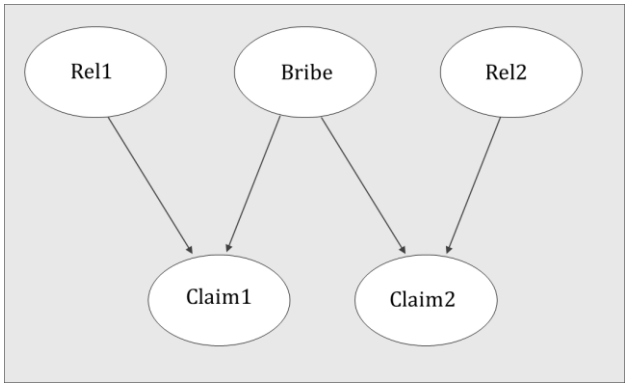

Figure 1: Bayesian Network of the political bribe vignette used in our study. Bribe = whether or not the politician took a bribe; Claim1 = prosecutor accuses politician of taking a bribe; Claim2 = another prosecutor corrects this claim; Rel1 = reliability of the prosecutor accusing the politician of taking a bribe; Rel2 = reliability of prosecutor correcting this claim.

Method

Participants

200 US participants (99 female, 100 male, 1 non-binary: mean age 39, range 18-76) were recruited from Prolific and paid them £9/hour for 13 minutes. The sample was split evenly between Democrats (100) and Republicans (100). An a priori power analysis suggested that to detect an effect of size F =.25 would require a minimum sample size of 159 participants, 53 per condition (a priori repeated-measures ANOVA; $\alpha = .05$; $1 - \beta = .8$; one group; two measurements; no nonsphericity correction).

Design

Participants were presented with a vignette about a politician running for re-election. Midway through, they encountered a claim from a prosecutor accusing the politician of taking a bribe, followed by a correction from a second prosecutor denying the claim. The vignette included three conditions, manipulating the politician's political affiliation (Republican, Democrat, or Unspecified). The study used a 3x2 within-subjects design, analysing how participants updated their belief in the bribery claim depending on their political alignment with the fictional politician and how they evaluated the reliability of the prosecutors. Further information on the design can be found on OSF.

Procedure

Participants were instructed to read a political vignette and answer related questions, requiring them to recollect parts of the story. Participants were first given background on the politician's campaign, followed by a prosecutor's claiming he took a bribe. Additional messages about the politician's family and events were presented before a final message from another prosecutor stating that the bribery did not occur. Participants read the messages at their own pace, but could not revisit previous ones. Afterward, they answered attention-check questions. After a brief description of the politician, participants prior beliefs were measured on a 0-100 slider scale, using the question: "How likely do you think it is that Henry Light took bribe money during his election campaign?". The slider was set by default at 50 and clicking on the slider was required to progress. The same question was asked after the bribery accusation by the first prosecutor and after the bribery correction by the second prosecutor. Source reliability was measured at three stages: before the vignette, after the bribery claim, and after the correction. Participants were asked to rate the reliability of the prosecutor at each stage, based on the information available at that time.

Crucially, following these questions at each stage participants were also asked a series of questions related to the conditional probabilities of the bribery claim under the supposition that the bribe was (or was not) taken and whether or not the prosecutor was reliable. These questions allowed us to parameterise the BN (shown in Figure 1) by eliciting values for the conditional probability tables for both Claim1 and Claim2. We could then use these individualised BNs to assess whether each participant updated their beliefs in line with the Bayesian model and also assess the influence of source reliability. Specifically, participants were asked to rate their probability in a claim being made by a prosecutor under four different conditions (reliable vs unreliable prosecutor, bribe vs no bribe). The questions followed the format:

"If Henry Light took/did not take the bribe and the prosecutor is reliable/unreliable, how likely is it that the prosecutor would report the bribery occurred?". Given the cognitive complexity of the questions, we preceded this measure with the following qualification: "In this section, we would like you to imagine hypothetical scenarios involving Henry Light and the prosecutor's testimony. For each scenario, we will ask you to estimate how likely it is that the prosecutor would report that the bribery occurred […] For each scenario, please rate how likely it is that the prosecutor would report that the bribery occurred, based on the combination of these factors. The scale goes from 0 (very unlikely) to 100 (very likely)." The questions were presented all on the same page to give participants the opportunity to reflect on them and adjust their measures where necessary.

At the end of the survey participants answered demographic questions such as gender, education and political affiliation. The latter was asked both on a continuous scale from 0-10:

"Where would you place yourself on this scale in terms of political orientation?" as well as a categorical question: "In general, what is your political affiliation?" to which the answer was "Republican", "Democrat" or "Other/Prefer not to say". This was to check that the categories used by Prolific to split our sample by political affiliation were not outdated.

Analyses

Participants were grouped into three categories of 'worldview' based on whether their political affiliation was aligned with the politician in the vignette (*congruent*) or not aligned (*incongruent*). Participants in the condition where the politician's party was undefined was the neutral condition.

We conducted a mixed design ANOVA to examine whether individuals' belief updating patterns varied between congruent and incongruent conditions as well as whether participants discounted corrected information. The same analysis was conducted on the reliability ratings.

For the Bayesian network model, for each participant, we used R package gRain to calculate the Bayesian-predicted posterior belief based on their prior beliefs and the conditional probabilities they provided (see osf for further details on fitting procedures). We did this both for the interim posterior after they read about the bribery claim as well as after the correction. We checked whether the posterior belief after correction was consistent with the Bayesian-predicted posterior derived from the participant's initial prior belief and their conditional probabilities. For each participant, we computed the difference between their initial posterior and the Bayesian-predicted posterior both after the bribery claim and after the correction. A statistically significant difference between the actual and predicted posteriors was interpreted as a deviation from Bayesian updating. A comparison between the baseline bribery beliefs and the post-correction beliefs was also used to test whether individuals show evidence of the continued influence effect (CIE): whether they continue to rely on corrected information or whether they effectively go back to baseline.

Building on this analysis we then compared these patterns of belief updating to those in the ideologically charged conditions. Specifically, we looked at whether participants are more likely to update their beliefs in a Bayesian manner when the evidence aligns with their worldview (i.e., when the politician does not share their political affiliation) and less likely when the evidence contradicts their worldview. We computed participants' posteriors after the claim and after the correction as above. We then used t-tests to examine the difference between the posteriors in each condition to our Bayesian predicted measures. Finally, we compared the differences between observed and predicted values across conditions to see whether the condition influences the divergence from the predicted model.

Results

Belief Updating

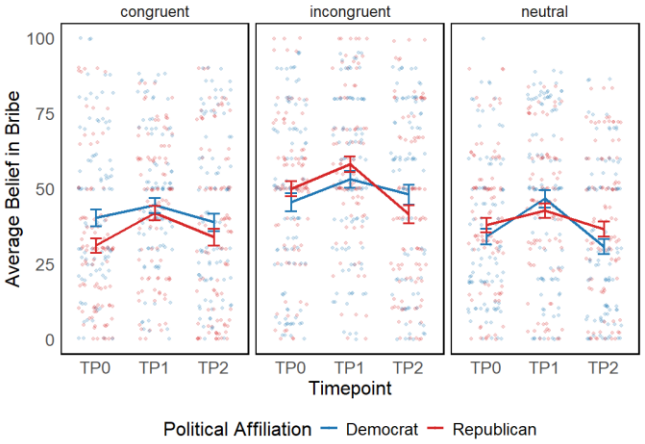

Figure 2: Belief in bribery across timepoints by political affiliation

Our results reveal two key patterns: (1) beliefs were higher in the incongruent condition, likely due to prior expectations about political malfeasance; (2) belief in bribery increased after the accusation (TP1) but returned to baseline after the correction (TP2) across all conditions, indicating effective belief updating (see Figure 2). We used a mixed-design ANOVA to examine the effects of worldview (congruent, incongruent, neutral), participant political affiliation (Republican or Democrat), and timepoint (prior to the claim, pre-correction, and post-correction) on beliefs about bribery occurrence. We found a significant effect of worldview, $F(2, 193) = 5.18$, $p = 0.006$, indicating that participants' beliefs were influenced by the congruence between their own political affiliation and the politician's. Timepoint also had a significant effect, $F(2, 386) = 16.34$, $p < 0.001$, showing that beliefs changed across time. The interaction between worldview and Timepoint was not significant, $F(4, 392) = 0.34$, $p = .849$, indicating that belief trajectories did not differ across worldview conditions over time. Political affiliation also had no significant effect on bribery belief, $F(1, 193) = 0.08$, $p = 0.77$, suggesting no difference in belief change between Republicans and Democrats. These results show that belief updating patterns differed significantly between congruent and incongruent conditions, contradicting the alternative hypothesis that participants update beliefs similarly, albeit with different priors. Crucially, however, across all conditions (incongruent: estimate = -7.89, $p = 0.018$; congruent: estimate = -7.46, $p = 0.022$; neutral: estimate = -8.65, $p < 0.01$), participants updated their beliefs when reading about the claim, and across all conditions they returned back to baseline with no statistically significant difference between the belief in the bribe prior to the claim and after the correction (incongruent: estimate = 2.95, $p = 0.64$; congruent: estimate = -0.64, $p = 0.98$; neutral: estimate = 2.31, $p = 0.75$). These findings suggest that participants, regardless of the worldview condition, responded with some

degree of updating to the claim but returned to their prior beliefs after the correction. Contrary to prior research we therefore find no evidence of the CIE or of continued reliance on misinformation due to politically motivated thinking.

Bayesian Network analysis

To further explore these findings, we compared participants' beliefs about the bribe across timepoints with Bayesian Network model predictions in the congruent and incongruent conditions, aiming to assess whether participants' belief updating aligned with the diagnosticity they assigned to the evidence (see Figure 3). A close alignment between actual and predicted values would suggest that participants updated their beliefs in a Bayesian, or 'rational,' manner—adjusting their priors based on the weight of the evidence provided.

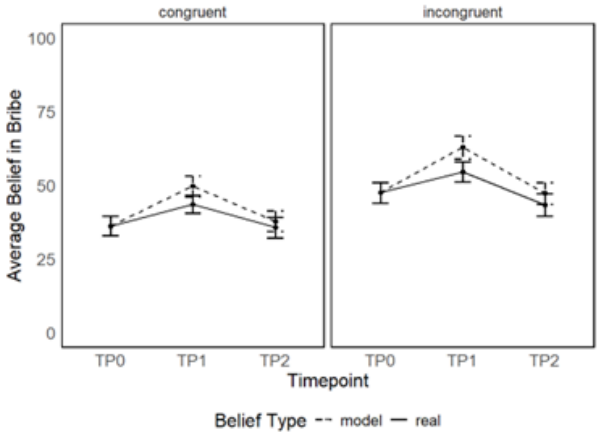

Figure 3: Comparison between actual and predicted values of average belief in bribe across conditions.

As previously outlined, t-tests were conducted to assess the difference between the Bayesian predicted posteriors and the measured posteriors after the claim (Timepoint 1) and after the correction (Timepoint 2) across different levels of worldview. At Timepoint 1, significant differences were found in all conditions. In both the congruent ($t(65) = -2.11$, $p = 0.039$), incongruent ($t(61) = -2.34$, $p = 0.022$) conditions the actual stated beliefs in the bribe were significantly lower than the model predictions. These results are in line with past research showing that belief updating tends to be conservative compared to model predictions (Kovach, 2021). In contrast, after the correction, the pattern shifted. For both the congruent ($t(65) = -0.90$, $p = 0.37$) and incongruent ($t(61) = -1.62$, $p = 0.11$) conditions there was no statistically significant difference between the actual and predicted values of belief. A Pearson's correlation analysis further supported these results showing a positive significant correlation between predicted and observed values $r(196)=.51$, 95% CI [.40, .61], $t(196) = 8.32$, $p<.001$. These findings suggest that although participants might be more conservative in their belief updating than predicted by the model, across both congruent and incongruent conditions participants effectively corrected their beliefs in the claim going back to baseline as dictated by the model.

Reliability

We examined changes in the perceived reliability of the source based on their claim and the experimental condition. Since the only manipulated factor in this experiment is the political affiliation of the politician, we compared the reliability ratings of the accuser and the corrector against a baseline reliability rating of a prosecutor, which was collected prior to the vignette. The goal is to determine whether the prosecutor making the bribery accusation is perceived as less reliable in conditions where the participant's political worldview aligns with the politician's affiliation. Additionally, we compared the perceived reliability between the accuser and the corrector within these conditions.

A mixed-design ANOVA was conducted to examine the effects of worldview (congruent, incongruent, neutral), timepoint (pre-claim, pre-correction, post-correction), and their interaction on reliability ratings. Prior to the vignette the question refers to the reliability of a prosecutor in general while the other two measurements refer to the reliability of the prosecutors making the claim and the correction. The analysis revealed a significant main effect of timepoint, $F(2, 392) = 32.66$, $p < 0.001$, with higher reliability ratings at TP0 and TP2 compared to TP1 (see Figure 4). The main effect of worldview was not significant, $F(2, 196) = 2.05$, $p = 0.13$ however, the interaction between worldview and timepoint was significant, $F(4, 392) = 2.44$, $p = 0.0467$, indicating that the pattern of reliability ratings varied across worldview conditions.

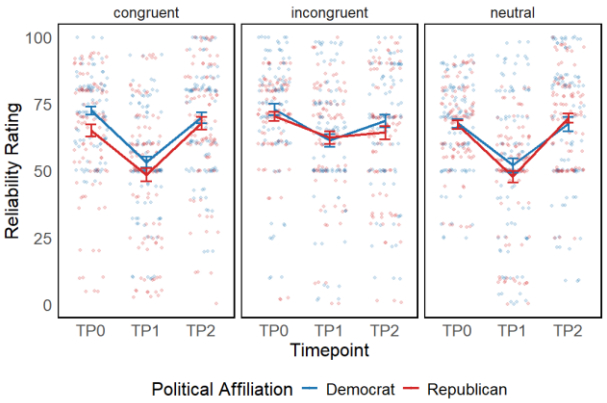

Figure 4: Prosecutor reliability across timepoints by political affiliation

Post-hoc analysis using estimated marginal means indicated that both Democrats and Republicans showed significant changes in their reliability ratings across timepoints. Participants rated the accuser's reliability significantly lower than the original prosecutor in all conditions: congruent (estimate = -17.93, $p < 0.001$), incongruent (estimate = -9.81, $p = 0.0089$), and neutral (estimate = -17.59, $p < 0.001$). Additionally, the reliability of the accuser was significantly lower than that of the corrector in the congruent (estimate = -17.99, $p = 0.001$) and neutral (estimate = -18.71, $p < 0.001$) conditions. However, no significant difference was observed between the accuser and the corrector in the incongruent

condition (estimate = -4.59, p = 0.51). For the corrector, reliability ratings returned to baseline across all conditions. No significant differences were observed between the original prosecutor and the corrector in any condition: congruent (estimate = -0.06, p = 0.99), incongruent (estimate = 5.22, p = 0.29), or neutral (estimate = -1.13, p = 0.94). These findings suggest that the accuser's reliability was significantly downgraded, particularly in the congruent condition, where the claim conflicted with the participants' worldview. A Welch Two-Sample t-test confirmed this, showing a significant difference in reliability ratings before and after the claim across conditions, t(369.31) = -3.02, p = 0.0027. The mean difference in the congruent condition (M = -17.91) was significantly larger than in the incongruent condition (M = -9.81), showing that participants downgraded the accuser's reliability more when the claim contradicted their views. In contrast, the decline in reliability was less pronounced when the accuser's claim was about a politician with opposing views. For the corrector, however, results suggest that the reliability of this prosecutor was restored and was no different to the reliability of the prosecutor rated prior the vignette. These results show that although individuals might downgrade the reliability of an individual member of an institution when that member's claims don't align with their own, this penalisation doesn't generalise to other members of the institution.

Discussion

The present results provide evidence of rational belief updating by participants. Across all conditions, participants initially increased their belief in the bribe after the claim but then fully corrected their beliefs, returning to baseline levels following the correction. Consistent with our hypothesis and previous studies (Sanna & Lagnado, 2024; Ecker et al., 2020), individuals effectively corrected their beliefs, showing no evidence of the continued influence effect. Surprisingly, and contrary to our expectations, these findings were consistent across both worldview-congruent and incongruent conditions. This aligns with previous research highlighting individuals' capacity for analytic thinking (Pennycook & Rand, 2019) and suggests that reliance on misinformation can be reduced, as participants were able to discount corrected information, even in politically charged contexts. This pattern was observed for both Democrat and Republican participants. Furthermore, when comparing these updates to Bayesian Network models, participants' belief revisions were consistent with Bayesian updating based on their individual priors, judged diagnosticity of the evidence and their perceived reliability of the source. Participants' updates qualitatively aligned with the model predictions albeit being more conservative, a trend often encountered in the literature (Kovach, 2021). Following the correction participants were generally aligned with model predictions, with no statistically significant difference between actual and predicted values across conditions. Once again, these findings were consistent across both worldview congruent and incongruent conditions. These findings not only show the potential for rational belief updating by individuals but also identify Bayesian Network models as effective methodologies for understanding belief updating. This work can be extended to focus more closely on how people evaluate the impact of evidence relative to Bayesian norms (Powell & Nair, 2023).

With regards to reliability, individuals downgraded the reliability of the prosecutor in the worldview congruent condition but not in the incongruent condition. This could be interpreted as a form of motivated reasoning and an attempt to discredit a source that attacks a member of the participant's ingroup. This also shows that regardless of the fictional nature of the vignette, participants actively showcased their political beliefs. However, it is important to note that in the congruent condition, the prosecutor's claim contradicted participants' low prior belief in the bribe. Given that participants rated the likelihood of the bribe as low, the claim was both surprising and inconsistent with their prior beliefs, which could reasonably explain the downgrade in reliability. The reliability of the corrector, however, did not show the same downgrade, even in the incongruent condition where the correction contradicted participants' political views. This suggests that while an individual may downgrade the reliability of a source, this effect does not necessarily extend to the broader institution. Specifically, the second prosecutor's reliability was recovered, indicating that the overall institutional credibility of prosecutors remained intact despite individual biases. This challenges the common assumption that the reputation of one individual can tarnish the entire institution. Furthermore, this finding highlights the potential for corrective actions by other members of the same institution to restore trust and improve reliability.

Given these results, it is important to consider the impact of the specific vignette on participants' responses, as real-world scenarios may involve more complex factors compared to fictional ones. Future research should explore the generalisability of these findings by testing them with diverse scenarios and real-world figures. Additionally, the sources in our study were generally perceived as highly reliable, but vignettes featuring less reliable sources might trigger stronger motivated reasoning. Finally, while participants demonstrated internal consistency in their belief updating, this does not necessarily equate to rationality, as they may still engage in inconsistent reasoning at various stages of the process. Future studies should broaden the scope of these findings by examining whether people adopt different models of the situation, for example by eliciting Bayesian network models from the participants.

Despite these limitations, this paper sheds light on crucial dynamics in belief updating processes. While individuals seem to have different updating patterns dependent on worldview congruence, in each condition, and regardless of political orientation, corrections seem to be generally effective at returning individuals to their baseline beliefs, an optimistic finding which can inform attempts at correcting misinformation in politically charged contexts.

References

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social media+ society*, *9*(1), 20563051221150412.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. Journal of Personality and Social Psychology, 39(6), 1037–1049. https://doi.org/10.1037/h0077720

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. Social science computer review, 37(1), 38-54.

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

Connor Desai, S. A., Pilditch, T. D., & Madsen, J. K. (2020). The rational continued influence of misinformation. Cognition, 205, 104453. https://doi.org/10.1016/j.cognition.2020.104453

Connor Desai, S., & Reimers, S. (2019). Comparing the use of open and closed questions for Web-based measures of the continued-influence effect. Behavior Research Methods, 51(3), 1426–1440. https://doi.org/10.3758/s13428-018-1066-z

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

Ecker, U. K. H., Lewandowsky, S., & Swire, B. (2020). The continued influence effect of misinformation: The role of political worldview. Journal of Applied Research in Memory and Cognition, 9(4), 592–603. https://doi.org/10.1016/j.jarmac.2020.08.005

Fenton, N., Neil, M., & Lagnado, D. A. (2013). A general structure for legal arguments about evidence using Bayesian networks. Cognitive science, 37(1), 61-102.

Gershman, S. J. (2019). How to never be wrong. Psychonomic Bulletin & Review, 26(1), 13–28. https://doi.org/10.3758/s13423-018-1488-8

Guillory, J. J., & Geraci, L. (2010). The persistence of inferences in memory for younger and older adults: Remembering facts and believing inferences. Psychonomic Bulletin & Review, 17(1), 73–81. https://doi.org/10.3758/PBR.17.1.73

Harris, A. J. L., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(5), 1366–1373. https://doi.org/10.1037/a0016567

Harris, A. J. L., Hahn, U., Madsen, J. K., & Hsu, A. S. (2016). The Appeal to Expert Opinion: Quantitative Support for a Bayesian Network Approach. Cognitive Science, 40(6), 1496–1533. https://doi.org/10.1111/cogs.12276

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. Social Cognition, 27(5), 733–763. https://doi.org/10.1521/soco.2009.27.5.733

Jern, A., Chang, K. K., & Kemp, C. (2014). Belief polarization is not always irrational. Psychological Review, 121(2), 206–224. https://doi.org/10.1037/a0035941

Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(6), 1420–1436. https://doi.org/10.1037/0278-7393.20.6.1420

Kovach, M. (2021). Conservative updating. arXiv preprint arXiv:2102.00152.

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

Lagnado, D. A. (2021). Explaining the evidence: How the mind investigates the world. Cambridge University Press.

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

Madsen, J. (2016). Trump supported it?! A Bayesian source credibility model applied to appeals to specific American presidential candidates' opinions. Cognitive Science. https://www.semanticscholar.org/paper/Trump-supported-it!-A-Bayesian-source-credibility-Madsen/ccdbba50f93e81fdcb88b94828d68a760568d367

Madsen, J. K., Hahn, U., & Pilditch, T. D. (2020). The impact of partial source dependence on belief and reliability revision. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(9), 1795–1805. https://doi.org/10.1037/xlm0000846

Merdes, C., von Sydow, M., & Hahn, U. (2021). Formal models of source reliability. Synthese, 198(23), 5773–5801. https://doi.org/10.1007/s11229-020-02595-2

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. Political Behavior, 32(3), 303–330. https://doi.org/10.1007/s11109-010-9112-2

O'Rear, A. E., & Radvansky, G. A. (2020). Failure to accept retractions: A contribution to the continued influence effect. Memory & Cognition, 48(1), 127–144. https://doi.org/10.3758/s13421-019-00967-9

Oyserman, D., & Dawson, A. (2020). Your fake news, our facts: Identity-based motivation shapes what we believe, share, and accept. In The psychology of fake news (pp. 173-195). Routledge.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers: San Francisco, CA.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 116(7), 2521–2526. https://doi.org/10.1073/pnas.1809789116

Pilgrim, C., Sanborn, A., Malthouse, E., & Hills, T. T. (2024). Confirmation bias emerges from an approximation to Bayesian reasoning. Cognition, 245, 105693. https://doi.org/10.1016/j.cognition.2023.105693

Powell, D., & Nair, S. (2023). Bayesian confirmation and commonsense notions of evidential strength. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 45, No. 45).

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

Ross, A. S., & Rivers, D. J. (2018). Discursive deflection: Accusation of "fake news" and the spread of mis-and disinformation in the tweets of President Trump. Social media+ society, 4(2), 2056305118776010.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32(5), 880–892. https://doi.org/10.1037/0022-3514.32.5.880

Sanna, G. A. and Lagnado, D. (Preprint published in 2024, July 27), Rational Belief Updating in the Face of Misinformation: The Role of Source Reliability. Available at SSRN: https://ssrn.com/abstract=4907730 or http://dx.doi.org/10.2139/ssrn.4907730

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

Thaler, M. (2024). The fake news effect: Experimentally identifying motivated reasoning using trust in news. American Economic Journal: Microeconomics, 16(2), 1-38.

Wilkes, A. L., & Leatherbarrow, M. (1988). Editing episodic memory following the identification of error. The Quarterly Journal of Experimental Psychology Section A, 40(2), 361–387. https://doi.org/10.1080/02724988843000168

Wilkes, A. L., & Reynolds, D. J. (1999). On certain limitations accompanying readers' interpretations of corrections in episodic text. The Quarterly Journal of Experimental Psychology Section A, 52(1), 165–183. https://doi.org/10.1080/713755808