BRIEF REPORT

Closed-minded cognition: Right-wing authoritarianism is negatively related to belief updating following prediction error

Alyssa H. Sinclair[1] · Matthew L. Stanley[1,2] · Paul Seli[2]

Abstract

When confronted with information that challenges our beliefs, we must often learn from error in order to successfully navigate the world. Past studies in reinforcement learning and educational psychology have linked *prediction error*, a measure of surprise, to successful learning from feedback. However, there are substantial individual differences in belief-updating success, and the psychological factors that influence belief updating remain unclear. Here, we identify a novel factor that may predict belief updating: right-wing authoritarianism (RWA), which is characterized by a desire for order, structure, and preservation of social norms. We hypothesized that because people who score high on RWA are motivated to preserve entrenched beliefs, they may often fail to successfully update their beliefs when confronted with new information. Using a novel paradigm, we challenged participants' false beliefs and misconceptions to elicit prediction error. In two studies, we found consistent evidence that high-RWA individuals were less successful at correcting their false beliefs. Relative to low-RWA individuals, high-RWA individuals were less likely to revise beliefs in response to prediction error. We argue that RWA is associated with a relatively closed-minded cognitive style that negatively influences belief updating.

Keywords Right-wing authoritarianism · Belief updating · Misinformation · Prediction error · Actively open-minded thinking

Learning from error is an essential part of adaptive memory processes. In the real world, we may read an inaccurate news story, make mistakes on an exam, or receive health-care information that necessitates making lifestyle changes. Recently, during the COVID-19 pandemic, Americans received conflicting health-care information that first discouraged, then later encouraged, wearing masks. Under these circumstances, we should update our beliefs in order to correct misinformation and learn from our mistakes.

Although learning from error is clearly important for success, humans are sometimes remarkably resistant to change. In fact, feedback that challenges an erroneous belief can paradoxically *strengthen* one's misconceptions (Ecker, Hogan, & Lewandowsky, 2017; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Nyhan & Reifler, 2010; Nyhan, Reifler, Richey, & Freed, 2014). Although research has indicated that there are individual differences in people's ability to successfully update their beliefs (Lewandowsky et al., 2012), there is a relative dearth of research on the psychological and ideological factors that might account for these individual differences. Here, we implicate a novel predictor of belief updating: right-wing authoritarianism (RWA). We investigate how RWA relates to underlying cognitive processes of learning from error, providing new insight into factors that influence belief-updating success.

✉ Alyssa H. Sinclair
     allie.sinclair@duke.edu

[1] Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

[2] Department of Psychology & Neuroscience, Duke University, Durham, NC, USA

Learning from error

Prior research has examined several systematic and cognitive factors that influence belief updating. Yet it remains unknown how individual differences in personality and ideology predict belief updating. The cognitive mechanisms of learning from error have been extensively studied across a broad range of

domains, including computational reinforcement learning (Rescorla & Wagner, 1972; Sutton & Barto, 1998), educational psychology (Butler, Fazio, & Marsh, 2011; Butterfield & Metcalfe, 2001; Metcalfe, 2017), and dopaminergic reward systems in the brain (Bayer & Glimcher, 2005; Schultz, Dayan, & Montague, 1997; Watabe-Uchida, Eshel, & Uchida, 2017). In reinforcement learning paradigms, the brain is thought to calculate a *prediction error* when people encounter surprising feedback; the strength of this error signal indicates the discrepancy between expectation and reality (Watabe-Uchida et al., 2017). Stronger prediction errors, which reflect greater surprise, tend to enhance learning and knowledge updating.

However, in some cases, we do *not* learn from corrective feedback. For instance, research on the *continued influence effect* has shown that misinformation can exert powerful, persistent effects on memory (Frenda, Nichols, & Loftus, 2011; Johnson & Seifert, 1994; Lewandowsky et al., 2012; Loftus, 2005). Even after misinformation is explicitly debunked, belief in the misinformation often persists (Southwell & Thorson, 2015; Thorson, 2016). Misinformation is particularly enduring and resistant to correction when it aligns with established beliefs or identities (Ecker & Ang, 2019; Ecker et al., 2017; Ecker, Lewandowsky, Fenton, & Martin, 2014). Moreover, repeating a falsehood during correction (e.g., "The news report that previously stated X was incorrect") makes successful belief revision *less likely*, because repetition eases the processing of the falsehood, and processing fluency serves as a heuristic for truth (Alter & Oppenheimer, 2009; Lewandowsky et al., 2012; cf. Meyer et al., 2015). Thus, retractions that challenge beliefs can paradoxically strengthen misconceptions over time. Contrary to predictions from reinforcement learning, surprising feedback does *not* always lead to belief updating. Although it is well-established that surprising feedback may not always lead to belief updating, it remains unclear which psychological factors may explain individual differences in belief updating.

Right-wing authoritarianism and cognitive styles

Right-wing authoritarianism (RWA) refers to a broad ideological framework for interacting with the world (Duckitt, 2001; Van Hiel, Cornelis, & Roets, 2007). RWA involves submission to authority figures, adherence to social norms and beliefs, and hostility towards those who challenge social norms (Altemeyer, 1998; Duckitt, 2001; Duckitt & Bizumic, 2013; Duckitt, Bizumic, Krauss, & Heled, 2010). People who score high on RWA tend to endorse structured societal systems that enforce uniformity, consistency, structure, and control, even through the use of coercion and restriction (Duckitt & Bizumic, 2013; Van Hiel et al., 2007). Therefore, we might

expect that in the context of belief updating, high-RWA individuals would prioritize order, consistency, and uniformity rather than truth. In the present study, we explore whether high-RWA individuals tend to be more closed-minded. If high-RWA individuals are particularly closed-minded when presented with new information, then those individuals may be more resistant to changing their beliefs in the face of surprising or contradictory evidence.

RWA has been primarily studied in relation to prejudice and beliefs about sociopolitical issues (Harnish, Bridges, & Gump, 2018; Meloen, 2019). However, there has been a recent surge of interest in the underlying cognitive styles that contribute to RWA opinions and behaviors. RWA has been linked to low openness-to-experience and denial of scientific evidence (e.g., climate change denial; Butler, 2000; Häkkinen & Akrami, 2014). A meta-analysis has also shown that RWA ideologies and prejudice are related to lower cognitive reasoning abilities (Onraet et al., 2015). Recent evidence supports the idea that a closed-minded cognitive style is the strongest predictor of RWA, above and beyond other cognitive styles and social cognition scales (Berggren, Akrami, Bergh, & Ekehammar, 2019). Critically, the foregoing traits are closely related to rigid cognitive styles with high intolerance of ambiguity. Right-wing radicalism and closed-mindedness contribute to binary, black-and-white thinking and ambiguity intolerance (Lauriola, Foschi, & Marchegiani, 2015). Belief updating requires receptivity to contradictory information and a willingness to accept uncertainty. It therefore stands to reason that elements of a closed-minded cognitive style may impair successful belief updating.

Open-mindedness (or "openness/intellect," as defined by the Big Five personality factors), on the other hand, appears to be beneficial for belief updating. Openness captures people's readiness to adapt their attitudes and behaviors when exposed to new ideas and situations (Digman, 1990). Similarly, actively open-minded thinking (AOT) is a cognitive style characterized by a tendency to value new evidence that contradicts existing beliefs and to carefully consider others' opinions (Voss, Perkins, & Segal, 1991; Haran, Ritov, & Mellers, 2013). AOT is positively correlated with openness, but also captures distinct cognitive tendencies that are relevant to flexible belief updating and information seeking (Baron, 2019; Haran et al., 2013; Stenhouse et al., 2018). Past studies have shown that openness is negatively related to RWA (J. C. Butler, 2000; Hodson & Sorrentino, 1999; Hotchin & West, 2018; Peterson, Smirles, & Wentworth, 1997). To our knowledge, no research has yet investigated the relationship between AOT and RWA. However, converging evidence suggests that open-minded versus closed-minded cognitive styles may influence cognitive processes of belief updating, shaping the way that we learn from error.

Present study

In this preregistered study, we adopted a multifaceted approach to understanding the relationship between RWA and belief updating. Drawing on findings and theory from reinforcement learning and educational psychology, we investigated how RWA is related to an underlying, apolitical cognitive process: learning from corrective feedback. To this end, we created a novel stimulus set of common misconceptions and urban myths that are widely believed. When participants indicated belief in these misconceptions, we provided feedback that the statement was a myth. Importantly, by challenging these beliefs, our goal was to elicit surprise and thus create opportunities for the participants to learn from error. We investigated how individual differences in RWA influenced belief updating in the face of surprising feedback. To foreshadow, in two samples, we found that RWA was negatively related to belief updating, and this effect interacted with prediction error and time-to-test.

Method

We collected data from two independent samples to test the replicability of our findings. With the exception of some minor methodological differences noted below (e.g., removal of some psychometric scales), the replication group (Sample 2) followed the same procedure as Sample 1. Results from the two samples are presented side-by-side in all figures. For Sample 1, the sample size, study design, and general analysis approach were preregistered (https://aspredicted.org/blind. php?x=dj6up9). All measures collected in Sample 1 are reported in the Supplemental Material. For Sample 2, we preregistered the study design, hypotheses, exclusion criteria, and planned analyses (https://osf.io/s6vyb). Our main findings from Sample 1 replicated in Sample 2. The study procedure and materials were approved by the Duke Campus Institutional Review Board. All participants provided informed consent before completing the study.

Participants

Sample 1 One-hundred and eighty-eight U.S. resident workers on Amazon's Mechanical Turk voluntarily participated in the study for monetary compensation. Participants were compensated with a total of $9 (Session 1: $5, Session 2: $4). Forty participants were excluded for failing attention checks (see below for more information), and an additional 25 participants did not return for the second session. After these exclusions, the final sample consisted of 123 participants (68.7% male, M_{age} = 36.79 years, SD_{age} = 11.62 years). The minimum target sample size (N = 100) was determined prior to data collection based on results from an exploratory pilot study: A power

analysis based on the pilot results (r = .36 Pearson correlation between AOT scores and belief updating) found that a sample size of 94 participants would yield 95% power (α = .05). Thus, we set the minimum sample size at 100 participants and recruited additional participants in anticipation of attrition.

Sample 2 Two-hundred and six U.S. resident workers on Amazon's Mechanical Turk voluntarily participated in the study for monetary compensation. In order to participate, workers must have completed at least 500 previous human intelligence tasks with an average approval rating ≥97%. Workers who had completed a previous version of the study were restricted from participating. Participants were compensated with a total of $5 (Session 1: $3, Session 2: $2). Total compensation was lower in Sample 2 because the duration of the study was reduced (several inconsequential individual differences measures were removed). Twenty-seven participants were excluded for failing attention checks, and an additional 24 participants failed to return for the second session. After these exclusions, the final sample size was 155 participants (52.7% male, M_{age} = 37.58 years, SD_{age} = 11.06 years). The minimum target sample size (N = 150) to achieve 80% power was determined prior to data collection by a power analysis based on the significant correlation (Pearson's r = −.20) between RWA scores and average belief-updating scores found in Sample 1. In anticipation of attrition and exclusions, we recruited additional subjects. Note that the power analysis for Sample 1 was based on the AOT effect because we did not measure RWA in the pilot sample; the sample size for Sample 2 was determined by the effect size of the RWA correlation in Sample 1. If the same power analysis criterion (AOT effect size) is applied to Sample 2, the sample size remains sufficient (>100 subjects).

Materials

We developed a novel set of stimuli to assess belief updating. The stimulus set consisted of 120 trivia statements (see Supplemental Material, Table S1), 80 of which described widely believed urban myths. Examples of these urban myths include, "Eating before swimming increases the risk of cramps"; "Adding salt to a pot of water makes it boil faster"; "Deoxygenated blood in your veins is blue"; and "Diamonds are formed when coal undergoes high pressure." All of these statements are commonly believed to be true, but are in fact false. The remaining 40 items of the stimulus set were truthful statements that addressed general knowledge (e.g., "The United States of America has 50 states"; "Spinach is high in iron"). We included these general-knowledge items among our urban myths so that participants would not learn that every statement was false. Analyses reported below focus only on questions that participants answered incorrectly, because the

goal of the study was to challenge and correct false beliefs. All surveys were administered using Qualtrics software.

Prior to beginning data collection, we divided the stimulus set into two test sets of 60 items (Test A and Test B). We counterbalanced the test items by running two groups of participants. Group A-B ($N_1 = 63$, $N_2 = 78$) completed Test A during Session 1, and then completed Test B after a 1-week delay (see below for more details on the delay). Group B-A ($N_1 = 60$, $N_2 = 77$) completed the tests in the reverse order.

Procedure

Feedback learning The study took place over two sessions, which were separated by a 1-week delay. During Session 1, participants completed a baseline feedback learning task (see Fig. 1). Participants were instructed that some statements would be true and that others would be urban myths, and that they would receive accurate feedback about each statement. Participants were also repeatedly instructed that they must not cheat by writing down or looking up the answers to the trivia statements. On each trial, a trivia statement was randomly selected and displayed, and the participant reported whether they believed that the statement was true or false. The participant also rated confidence in their reported belief on a scale from 0 (*not at all confident*) to 100 (*very confident*). These initial confidence ratings are critical for our later analyses because they serve as an indirect measure of belief strength. After providing confidence ratings, the participant viewed the statement again and was provided feedback about whether the statement was actually true or false. This process repeated

until the participant viewed all 120 statements with feedback. The task was self-paced, but if a participant attempted to advance to the next question too quickly (<3 s), they were blocked from advancing and warned that they must read carefully to provide accurate answers.

Personality assessments After completing the feedback learning task, participants completed several personality measures, including the short Right-Wing Authoritarianism scale (RWA; 1 = *strongly disagree* to 5 = *strongly agree*; Zakrisson, 2005) and the 7-item version of the Actively-Open Minded Thinking scale (AOT; 1 = *completely disagree* to 5 = *completely agree*; Haran et al., 2013). The RWA and AOT scales included the following example items, respectively: "Our country needs a powerful leader, in order to destroy the radical and immoral currents prevailing in society today" (RWA); "People should take into consideration evidence that goes against conclusions they favor" (AOT). Additionally, we collected demographic information about social/economic conservatism, age, gender, education, income, ethnicity, and history of psychiatric disorders. For the sake of brevity, here we focus on RWA and AOT, but descriptive statistics for all preregistered personality measures are reported in the Supplemental Material (Table S3). In Sample 2, we updated the AOT scale with a more recent 10-item version (1 = *completely disagree* to 5 = *completely agree*; Baron, 2019). We also added a short Social Dominance Orientation (SDO) scale (1 = *strongly oppose* to 7 = *strongly favor*; Ho et al., 2015), because this construct is closely related to RWA and may also be related to belief updating (Pratto,
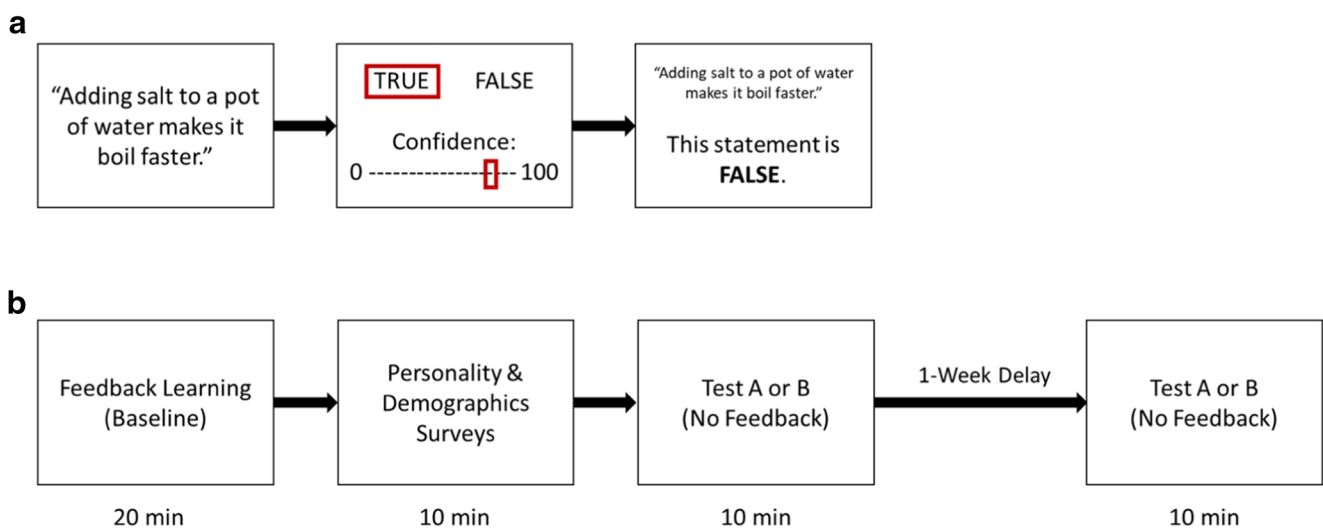

Fig. 1 Overview of the experimental procedure. **a** Example trial from the feedback learning task. Participants viewed 120 trivia statements, 80 of which were urban myths that are commonly believed. On each trial, the participant reported whether they believed that the statement was true or false, and then rated their confidence on a scale from 0 to 100. Each trial concluded with feedback about the correct answer. **b** Overview of the

two-session procedure. During Session 1, participants completed the baseline feedback learning task, personality and demographics surveys, and a test on half of the stimulus items. After a 1-week delay, participants completed a second test on the remaining half of the items. The order of the two tests was counterbalanced

Sidanius, Stallworth, & Malle, 1994). The revised AOT scale and the SDO scale included the following example items, respectively: "When faced with a puzzling question, we should try to consider more than one possible answer before reaching a conclusion" (AOT); "An ideal society requires some groups to be on the top and others to be on the bottom" (SDO). Descriptive statistics and Pearson product-moment correlations for all measures collected are provided in Table 3 and Tables S2a–b in the Supplemental Material.

Immediate and delayed tests The last component of Session 1 was a test on half of the trivia statements (counterbalancing methods are described in the Materials section). Participants were instructed to answer on the basis of the feedback they previously received during the learning phase of the study. The format of the test was identical to the initial learning task, except that the participants did not receive feedback about whether each statement was actually true or false. As before, participants viewed trivia statements, made a true-or-false judgement, and rated confidence on a scale from 0 to 100. After a 1-week delay, participants completed a second test on the previously untested half of the stimulus set.

Attention checks We included several attention checks throughout the experiment. The feedback learning task and the two tests included a trial on which participants were presented with the following text: "If you are paying attention, choose TRUE and set your confidence rating at 70." If participants failed any instance of this attention check question, we excluded all of their data from our analyses. We accepted confidence ratings between 65 and 75 to allow for minor errors with the sliding scale. Additionally, at the end of the personality surveys, participants responded to a question that stated, "Do you feel that you paid attention, avoided distractions, and took this survey seriously?" Only participants who responded "yes" to this question were included in analyses. We assured participants that their responses to this question would not affect their payment or eligibility for future studies.

Discouraging cheating Although it is difficult to entirely prevent cheating in an online study, we took several steps to deincentivize and identify cheating. In order to deincentivize cheating, during the consent process and instructions we repeatedly emphasized that accuracy on the trivia task would not determine participant compensation or job approval. This fixed payment schedule made cheating costly rather than beneficial; spending additional time and effort to look up the answers would not increase payment.

After the trivia task, participants were also asked to read the following honor code and self-report cheating: "Please be honest when answering the following question. Your answer will NOT affect your payment or eligibility for future studies. The study you have just participated in is a psychological study aimed at understanding human cognition and behavior. Psychological research depends on participants like you. Your responses to surveys like this one are an incredibly valuable source of data for researchers. It is therefore crucial for research that participants pay attention, avoid distractions, and take all study tasks seriously (even when they might seem silly). *Do you feel that you paid attention, avoided Googling the answers, and took this survey seriously?"* Participants who self-reported cheating or inattention were paid in full, but all of their data were excluded from analysis. These anticheating measures were repeated during every instance of the trivia task (learning and test).

Lastly, we identified trials wherein participants spent an excessive amount of time (>9 s, above the 80th percentile for trial duration) before advancing to the next trivia statement. Because these slow trials may indicate distraction or cheating (searching for answers or explanations online), we excluded these trials from analysis. Note that in the Sample 1 preregistration, we specified our intention to use a tab-monitoring tool to exclude trials where cheating may have occurred; due to technical difficulties with the tool, we adopted the duration exclusion rule instead. This duration rule was specified in the Sample 2 preregistration.

Statistical modeling Using R (Version 3.6), we constructed generalized logistic regression models with the *lme4* package (Bates, Mächler, Bolker, & Walker, 2014) and bound optimization by quadratic approximation (BOBYQA controller). We included crossed random effects to capture by-subject and by-item variability, with random slopes (for testing session and prediction error) and random intercepts. Continuous predictor variables were scaled and mean centered. All models converged successfully. Significance was assessed with two-sided Type III Wald tests (z stats and p values provided in-text throughout the Results section). Plots were produced with the *ggplot2* package (Wickham & Winston, 2019).

Given the nature of our stimuli, the distribution of trials by confidence ratings was skewed (see Fig. S1 in the Supplemental Material). The majority of the stimuli were endorsed with very high confidence (90–100), whereas few items failed to address a prior belief. The preregistration for Sample 2 specified our expectations about this skewed distribution, and stated that the prediction error measure would be binned or transformed as necessary. We binned the prediction error measure into five levels of confidence ratings from incorrect trials during the feedback learning phase (1 = [0–20], 2 = [20–40], 3 = [40–60], 4 = [60–80], 5 = [80–100]). This binning strategy optimizes interpretability for visualization, but produces an unequal number of observations per bin. We also tested the same statistical models and produced the corresponding figures with an alternate binning scheme that equated the number of observations per bin. These alternate-bin results are provided in the Supplemental Material (see

Tables S3, S4). Critically, the statistical significance of our primary findings was not affected by the choice of binning strategy.

Results

Challenging false beliefs to elicit prediction error

Our novel stimulus set effectively challenged existing beliefs. As intended, participants performed poorly at baseline, demonstrating many false beliefs. We conducted paired t tests to compare average accuracy across testing sessions. In Sample 1, overall percentage accuracy was low at baseline ($M = 0.59$, $SD = 0.14$), but significantly higher for both the immediate test ($M = 0.87$, $SD = 0.16$), $t(128) = 20.38$, $p < .001$, 95% CI [0.31, 0.26], and the delayed test ($M = 0.80$, $SD = 0.16$), $t(128) = 15.72$, $p < .001$, 95% CI [0.23, 0.17]. Similarly, in Sample 2, accuracy was low at baseline ($M = 0.59$, $SD = 0.12$), but significantly higher for both the immediate test ($M = 0.87$, $SD = 0.15$), $t(153) = 23.67$, $p < .001$, 95% CI [0.32, 0.27], and the delayed test ($M = 0.75$, $SD = 0.20$), $t(153) = 10.50$, $p < .001$, 95% CI [0.20, 0.13]. The stimulus set also produced a balanced set of responses at baseline; averaging across all trials, participants endorsed 56% of statements as "true" and 44% of statements as "false" in both samples. Overall, participants successfully learned from feedback and updated beliefs.

For the following analyses, we only analyzed the stimuli that participants answered incorrectly at baseline. We used the participant's confidence rating during the baseline learning task as a measure of prediction error (i.e., surprise). Confidence ratings during learning reflect the strength of false beliefs immediately before corrective feedback was provided. Confidence judgements have been previously used to quantify surprise in studies of learning from error (Metcalfe, 2017; Pine, Sadeh, Ben-Yakov, Dudai, & Mendelsohn, 2018). Here, we use confidence as a proxy for surprise, as previously used in a study of prediction error and knowledge updating (Pine et al., 2018). However, it is important to note that although this measure is inspired by ideas from the reinforcement learning framework, it is not a direct or model-derived error measure.

Right-wing authoritarianism is negatively related to belief updating

Next, we used generalized logistic mixed-effects regression to test whether individual differences were related to belief updating on a trial-wise basis. Descriptive statistics for parameter estimates are provided in Tables 1 and 2. The outcome measure in each model was *belief updating*, a binary variable that indicated whether a participant successfully corrected a false belief (1) or reproduced the same error at test (0). First, we tested each of the individual-differences measures in

Table 1 Logistic regression estimates for the effect of RWA on belief updating

Predictors	Odds ratios	SE	CI	p
Belief updating: Sample 1				
(Intercept)	7.19***	0.22	[4.65, 11.12]	**<.001**
RWA	0.61**	0.18	[0.43, 0.87]	**.006**
Prediction error	1.03	0.04	[0.95, 1.12]	.415
Session	0.25***	0.12	[0.20, 0.32]	**<.001**
RWA × Prediction Error	1.03	0.04	[0.95, 1.12]	.474
RWA × Prediction Error × Session	1.06*	0.03	[1.01, 1.12]	**.024**
Observations	4,368			
Marginal R^2	.144			
Belief updating: Sample 2				
(Intercept)	10.19***	0.23	[6.46, 16.06]	**<.001**
RWA	0.64*	0.18	[0.44, 0.91]	**.015**
Prediction error	0.94	0.05	[0.86, 1.03]	.209
Session	0.24***	0.13	[0.19, 0.31]	**<.001**
RWA × Prediction Error	0.96	0.05	[0.88, 1.05]	.401
RWA × Prediction Error × Session	1.08**	0.03	[1.02, 1.14]	**.010**
Observations	4,867			
Marginal R^2	0.126			

Boldface numbers indicate statistically significant parameters

*$p < .05$. **$p < .01$. ***$p < .001$

Table 2 Linear regression estimates for the effect of RWA on confidence change

Predictors	Change in confidence (Sample 1)				
	Estimate	SE	CI	p	df
(Intercept)	7.64***	1.67	[4.36, 10.92]	**<.001**	200.11
RWA	—−1.60	1.14	[−3.83, 0.63]	.162	102.47
Test accuracy	10.85***	1.53	[7.86, 13.85]	**<.001**	262.65
Session	−4.45*	1.86	[−8.09, −0.81]	**.017**	411.47
RWA × Test Accuracy	−2.57*	1.12	[−4.76, −0.38]	**.024**	104.71
Session × Test Accuracy	−5.47**	1.77	[−8.95, −1.99]	**.002**	2,885.50
Observations	4,611				
Marginal R^2 / Conditional R^2	0.065 / 0.308				
Predictors	Change in confidence (Sample 2)				
	Estimate	SE	CI	p	
(Intercept)	5.19**	1.59	[2.08, 8.31]	**.001**	
RWA	0.92	1.06	[−1.16, 3.01]	.387	
Test accuracy	13.19***	1.55	[10.16, 16.22]	**<.001**	
Session	−2.60	1.64	[−5.81, 0.62]	.114	
RWA × Test Accuracy	−4.81***	1.16	[−7.08, −2.53]	**<.001**	
Session × Test Accuracy	−5.74***	1.68	[−9.03, −2.45]	**.001**	
Observations	5,417				
Marginal R^2 / Conditional R^2	0.062 / 0.305				

Boldface numbers indicate statistically significant parameters

*$p < .05$. **$p < .01$. ***$p < .001$

separate models, along with interaction terms for prediction error and testing session.

In Sample 1, RWA was significantly negatively related to belief updating ($z = -2.73$, $p = .006$; see Table 1, top), indicating that individuals who scored higher on RWA tended to

be less successful at correcting erroneous beliefs after feedback (see Fig. 2). We also found a significant three-way interaction between RWA and prediction error ($z = 2.25$, $p = .024$; see Fig. 3). On the same-day test, high-RWA individuals were markedly worse than low-RWA individuals at belief updating
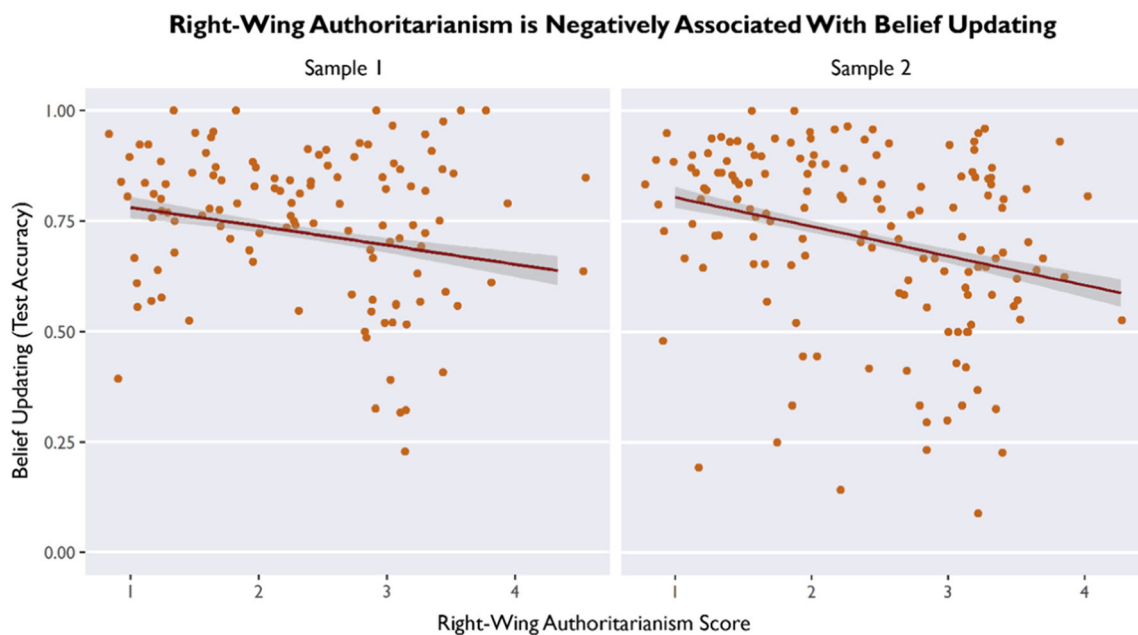

Fig. 2 RWA scores were negatively correlated with belief updating in both samples. Points depict subject averages, jittered slightly for visualization. Shaded band depicts 95% confidence interval for the line of best fit
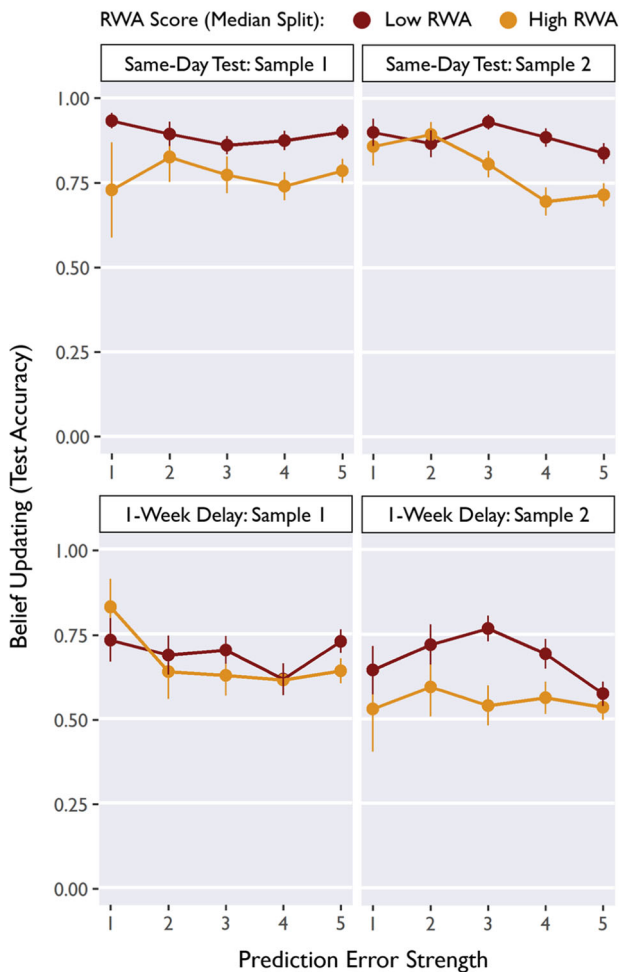
Fig. 3 Average belief updating scores across levels of prediction error. Lines depict a median split (for visualization purposes only) of low-RWA and high-RWA participants, and rows depict results from the same-day test (top) and 1-week delay test (bottom). Columns separate results by sample. Error bars depict standard error of the mean

for trials that had elicited moderate-to-strong prediction errors (Levels 3–5). High-RWA individuals were less likely to learn from feedback when it challenged a strongly-held belief. In contrast, low-RWA individuals were more successful at belief updating regardless of the level of prediction error. After a 1-week delay, high-RWA individuals demonstrated low belief updating scores across all levels of prediction error. There was also a significant main effect of testing session ($z = -11.54$, $p < .001$), whereby accuracy was lower after the 1-week delay. All of these findings replicated in Sample 2 (see Table 1, bottom). RWA was significantly negatively associated with belief updating, $z = -2.45$, $p = .015$. There was a significant three-way interaction among RWA scores, prediction error, and testing session ($z = 2.58$, $p = .01$). There was also a significant main effect of testing session ($z = -11.3$, $p < .001$).

Next, we conducted exploratory analyses to test whether RWA was related to confidence ratings. RWA was significantly positively correlated with average confidence ratings during the initial learning phase (Sample 1: Pearson's $r = .37$, $p < .001$; Sample 2: $r = .30$, $p < .001$). Despite high-RWA individuals reporting greater baseline confidence, we also found that RWA was negatively related to baseline accuracy (Sample 1: Pearson's $r = -0.24$, $p = .006$; Sample 2: $r = -0.27$, $p < .001$). Lastly, we calculated a difference score by subtracting the confidence rating at test from the confidence rating during learning. Using linear mixed-effects regression (see Table 2), we found a significant interaction between RWA scores and test accuracy predicting change in confidence (Sample 1: $t = -2.32$, $p = .022$; Sample 2: $t = -4.12$, $p < .001$). Low-RWA individuals showed a large increase in confidence when false beliefs were successfully corrected, but confidence was lower for high-RWA individuals (see Fig. 4). This suggests that even after learning from feedback, high-RWA individuals do not update beliefs as strongly.

Overall, in both samples, we found that high-RWA individuals were less successful at belief updating. High-RWA individuals were particularly resistant to correcting strongly held false beliefs when feedback elicited a large prediction error. This effect was strongest during the same-day test; after a 1-week delay, high-RWA individuals tended to show lower belief updating across all levels of prediction error. High-RWA individuals were also less confident in false beliefs that were corrected, suggesting that belief updating was weaker.

Actively open-minded thinking, conservatism, and social dominance orientation

Next, we tested several other measures that may contribute to understanding the relationship between RWA and belief updating. Pearson correlations among measures are provided in Table 3. In contrast to RWA, in Sample 1 we found that AOT was strongly positively associated with belief updating ($z = 3.80$, $p < .001$; see Table 4, top), with higher AOT scores predicting greater belief updating success (see Fig. 5). Comparable to RWA, but in the opposite direction, there was a significant three-way interaction among AOT scores, prediction error, and testing session ($z = -3.92$, $p < .001$). There was also a main effect of testing session ($z = -11.96$, $p < .001$) on belief updating. Most of these findings replicated in Sample 2 (see Table 4, bottom). AOT was significantly positively related to belief updating ($z = 2.98$, $p = .003$). There was a significant main effect of testing session ($z = -10.99$, $p < .001$). Additionally, there was a trending main effect of prediction error on belief updating ($z = -1.82$, $p = .069$). However, unlike in Sample 1, the three-way interaction among AOT, prediction error, and testing session was not significant. We also found that AOT was positively correlated with baseline accuracy in both samples (Sample 1: Pearson's r

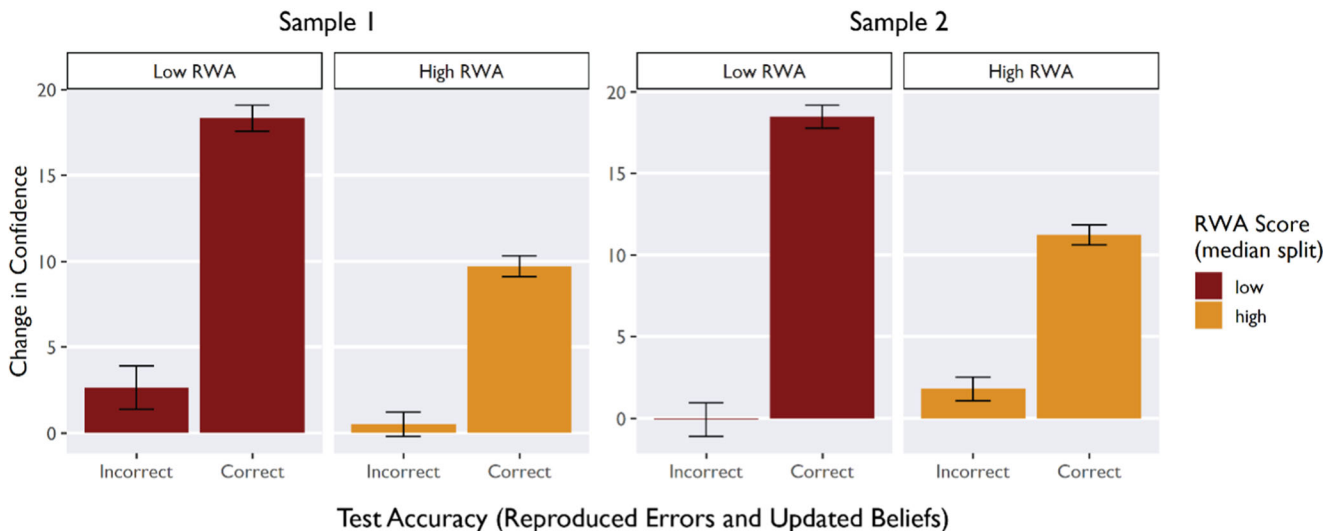## RWA is Negatively Related to Confidence in Updated Beliefs

Fig. 4 Average change in confidence for false beliefs that were correctly updated ("correct") or reproduced ("incorrect") at test. Relative to low-RWA individuals, high-RWA individuals were less confident in updated beliefs, suggesting that beliefs were not updated as strongly

= .34, $p < .001$; Sample 2: $r = .19$, $p = .017$). Overall, we found that AOT was positively associated with belief updating.

Lastly, we tested whether our belief-updating effects could be explained by conservatism and/or SDO. We found that neither conservatism (Sample 1: $z = -1.32$, $p = .187$; Sample 2: $z = 0.42$, $p = .673$) nor SDO (Sample 2: $z = -0.78$, $p = .436$)

were significantly related to belief updating. (The SDO analysis was restricted to Sample 2 because this measure was not collected in Sample 1.) Parameter estimates for these models are provided in Tables S5 and S6 in the Supplemental Material. Overall, neither conservatism nor SDO predicted belief updating.

Table 3 Descriptive statistics and correlations among individual differences measures

Sample 1: Means, standard deviations, and Pearson correlations with confidence intervals						
Variable		M	SD	1	2	3
1. Belief updating		0.75	0.18			
2. RWA		2.35	0.87	−.19*		
				[−.36, −.01]		
3. AOT		5.39	1.06	.47**	−.62**	
				[.32, .60]	[−.72, −.50]	
4. Conservatism		2.67	1.25	−.07	.57**	−.32**
				[−.25, .11]	[.44, .68]	[−.47, −.15]
Sample 2: Means, standard deviations, and Pearson correlations with confidence intervals						
Variable	M	SD	1	2	3	4
1. Belief updating	0.73	0.20				
2. RWA	2.34	0.83	−.32**			
			[−.45, −.17]			
3. AOT	4.41	0.83	.23**	−.34**		
			[.07, .37]	[−.47, −.19]		
4. SDO	1.76	1.81	−.01	.10	−.45**	
			[−.17, .15]	[−.06, .26]	[−.57, −.32]	
5. Conservatism	2.64	1.15	−.06	.60**	−.08	.03
			[−.22, .10]	[.49, .69]	[−.24, .07]	[−.13, .18]

Note. M and *SD* are used to represent mean and standard deviation, respectively. RWA = right-wing authoritarianism; AOT = actively open-minded thinking; SDO = social dominance orientation. Values in square brackets indicate the 95% confidence interval for each correlation. *$p < .05$. **$p < .01$

Table 4 Logistic regression estimates for the effect of AOT on belief updating

Predictors	Odds ratios	SE	CI	p
	Belief updating: Sample 1			
(Intercept)	6.52***	0.22	[4.22, 10.05]	**<.001**
AOT	2.08***	0.19	[1.43, 3.04]	**<.001**
Prediction error	1.06	0.04	[0.97, 1.15]	.217
Session	0.25***	0.11	[0.20, 0.32]	**<.001**
AOT × Prediction Error	1.00	0.04	[0.92, 1.09]	.971
AOT × Prediction Error × Session	0.91***	0.02	[0.87, 0.96]	**<.001**
Observations	4,368			
Marginal R^2	0.189			
	Belief updating: Sample 2			
(Intercept)	10.88***	0.23	[6.99, 16.94]	**<.001**
AOT	1.73**	0.18	[1.21, 2.49]	**.003**
Prediction Error	0.92	0.05	[0.84, 1.01]	.069
Session	0.25***	0.13	[0.19, 0.32]	**<.001**
AOT × Prediction Error	0.95	0.04	[0.87, 1.04]	.264
AOT × Prediction Error × Session	0.99	0.03	[0.94, 1.04]	.673
Observations	4,867			
Marginal R^2	0.104			

Boldface numbers indicate statistically significant parameters

AOT = actively open-minded thinking. *$p < .05$. **$p < .01$. ***$p < .001$

Discussion

Here, we investigated how right-wing authoritarianism (RWA) relates to belief updating. We challenged existing beliefs by testing participants on urban myths that are commonly endorsed as true. Critically, RWA scores were negatively related to belief updating: high-RWA individuals demonstrated an altered pattern of learning from error. Even for stimuli that did not challenge politically charged beliefs, high-RWA individuals demonstrated relatively poor belief updating. For the
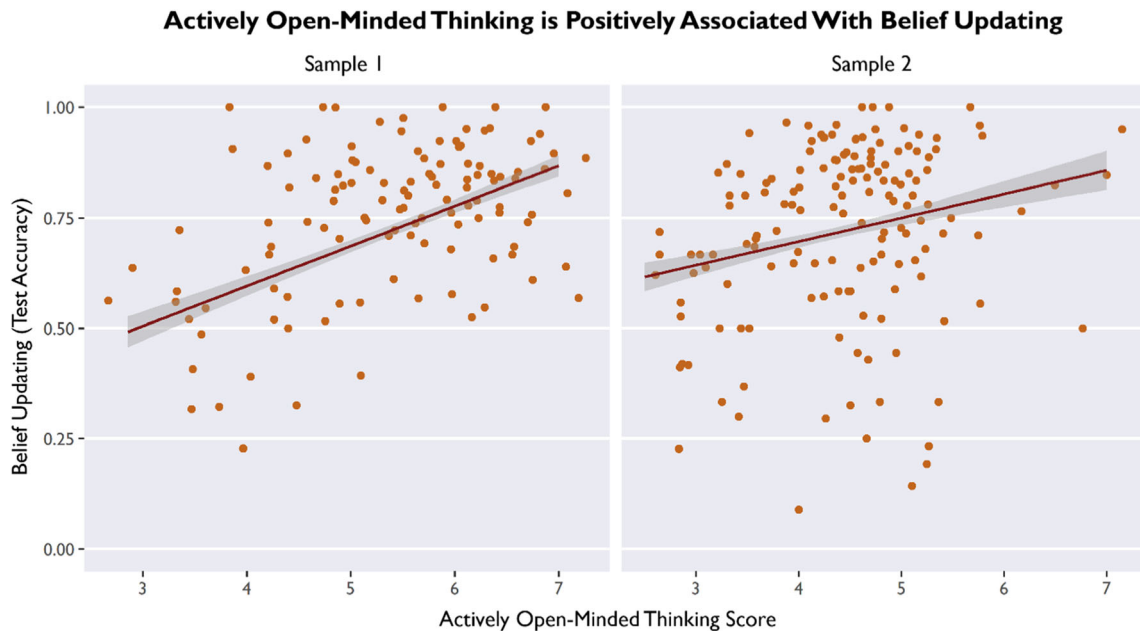

Fig. 5 Individual differences in actively open-minded thinking (AOT) predict belief updating success. AOT was positively correlated with belief updating in both samples. Points depict subject averages, jittered for visualization. Shaded band depicts 95% confidence interval for the line of best fit

first time, we also show that AOT is negatively related to RWA and positively related to belief updating. These complementary findings support the idea that a closed-minded cognitive style underlies authoritarian reasoning. Moreover, these individual-differences measures interacted with prediction error, a measure of surprise, and the delay to test. Overall, we show that individual differences in personality predict belief updating after surprising feedback.

Relating closed-minded thinking to cognitive processes of belief updating

We found that individual differences in personality were related to belief-updating success, and the extent to which prediction error governed learning. We defined prediction error as self-reported confidence in a false belief immediately before corrective feedback. If a participant strongly believed in a statement (e.g., reporting 90% confidence that the myth was accurate), then receiving feedback that challenged the belief would be very surprising. We found that RWA was negatively related to belief-updating success. Interestingly, RWA scores interacted with prediction error and testing session to predict belief updating: High-RWA participants tended to have lower belief-updating scores than did low-RWA individuals, especially at moderate-to-high levels of prediction error. This biased pattern of belief updating was most evident during the same-day test; after a 1-week delay, high-RWA individuals exhibited low belief-updating scores regardless of prediction error. We argue that the negative association between RWA and belief updating may reflect the desire for norm preservation and black-and-white thinking that are hallmarks of authoritarian reasoning.

Furthermore, we found that AOT was positively associated with belief updating. RWA and AOT were negatively correlated and inversely related to belief updating, supporting the idea that cognitive biases in authoritarianism stem from a closed-minded cognitive style (Berggren et al., 2019). Overall, we characterize subject-level and trial-level factors that shape how people learn from surprising feedback when established beliefs are challenged. Understanding how cognitive styles and the strength of beliefs interact to influence receptivity to feedback may guide future research on interventions to combat the dissemination of misinformation.

It is important to note that belief updating and motivated cognition likely reflect underlying cognitive tendencies that can span political boundaries. Radical beliefs have been linked to impaired belief updating and lower metacognitive sensitivity; on both ends of the political spectrum, radicals tend to be less sensitive to evidence that corrects mistakes (Rollwage, Dolan, & Fleming, 2018). Likewise, political liberals and conservatives appear to be equally likely to reject evidence that contradicts prior beliefs (Kraft, Lodge, & Taber, 2015; Nisbet, Cooper, & Garrett, 2015; Washburn & Skitka,

2018). In the present study, we found that neither conservatism nor SDO scores were related to belief updating. RWA measures may therefore capture a cognitive style that is uniquely related to belief updating. However, radicalism and biased information processing likely influence belief updating in other ideological groups. By investigating the cognitive mechanisms of belief updating, we aim to cast light on the common factors that may underly radicalization across political boundaries.

Future directions

Past studies have shown that stronger prediction errors are more likely to lead to knowledge updating (Metcalfe, 2017; Pine et al., 2018; Sutton & Barto, 1998). However, here we found that after a 1-week delay, belief updating was not very successful for trials that elicited very strong prediction errors. This accords with past evidence that high-confidence errors may be initially corrected after feedback, but can return after a delay (A. C. Butler et al., 2011). Importantly, in our paradigm, we challenged false beliefs rather than using difficult general-knowledge questions as in previous studies. Belief updating requires memory systems, as evidenced by the decline in accuracy after a delay to test. However, beliefs may be informed by affective or experiential elements that interact with memory processes. Here, we show that strong false beliefs may be more likely to resurface over time, even after surprising feedback. Future work may test whether there is a dissociation between revision of beliefs and general knowledge. Additionally, future studies may test prediction error with greater specificity by using a paradigm that derives error measures from computational models, as in reinforcement learning tasks.

To capture a range of prediction error, we used stimuli that probed moderate, everyday misconceptions. However, it is unclear whether our findings will generalize to strong beliefs in other domains. Because surprising feedback likely influences beliefs about highly contentious topics (e.g., climate change, vaccination) differently, future research should examine how prediction error influences strong, polarized beliefs. Our urban myths offer a novel way to investigate underlying cognitive mechanisms by targeting beliefs that are intermediate between impartial general knowledge and highly contentious issues. Here, we show that RWA is related to distinct a pattern of belief updating, even for apolitical stimuli.

In the present study, we limited our feedback to true/false statements. However, educational psychology studies have shown that providing detailed explanations during corrective feedback can improve knowledge updating (Butler, Godbole, & Marsh, 2013; Harbour, Evanovich, Sweigart, & Hughes, 2015; Metcalfe, 2017). On the other hand, providing additional information could fuel motivated reasoning. It may be fruitful for future research to explore whether RWA interacts with

prediction error during other cognitive tasks that involve information seeking or evaluation of evidence.

Lastly, it is also possible that RWA may be related to perceptions of source credibility. If high-RWA individuals are more likely to discredit sources that are unknown or believed to be untrustworthy, then feedback may be less effective. In the present study, we did not manipulate information about the source of the feedback. However, this information may have been implicitly provided by our university affiliation, which can suggest a liberal bias. Past work in social psychology supports the idea that belief updating is weighted by the credibility of the information source (e.g., Birnbaum & Mellers, 1983). Future studies may systematically manipulate source credentials and accuracy in order to test how belief updating, and the role of prediction error, differs depending on source perceptions.

Concluding remarks

Overall, we implicate RWA as a novel factor that relates to belief updating, and explore how cognitive styles may support learning from error. The inverse effects of RWA and AOT support the idea that a spectrum of closed-minded to open-minded cognitive styles influences belief updating. The present research on belief updating has broad implications for political polarization and the spread of misinformation. Misleading or incorrect information can persist in memory and continue to influence opinions and actions (Ecker et al., 2014; Johnson & Seifert, 1994). Strikingly, misinformation persists even after the source is discredited or the participant is reexposed to the correct information (Ecker et al., 2017). Misinformation in the media can exert pervasive effects on society, including inaccurate political news (Ecker & Ang, 2019), distortion of scientific evidence (Scheufele & Krause, 2019), climate change denial (Häkkinen & Akrami, 2014), and vaccine skepticism (Pluviano, Watt, & Della Sala, 2017). Understanding how prediction error, personality, and cognitive styles interact to support belief updating is crucial for counteracting polarized discourse and the spread of misinformation in the media.

Open practices The study reported here was pre-registered (Sample 1: https://aspredicted.org/blind.php?x = dj6up9; Sample 2: https://osf.io/s6vyb). Key findings replicated in Sample 2. Data sets from both samples and analysis code are provided in the Open Science Framework repository (https://osf.io/a85rq/).

References

Altemeyer, B. (1998). The other "authoritarian personality". *Advances in Experimental Social Psychology*, *30*(C), 47–92. doi:https://doi.org/10.1016/S0065-2601(08)60382-2

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. doi:https://doi.org/10.1177/1088868309341564

Baron, J. (2019). Actively open-minded thinking in politics. *Cognition*, *188*, 8–18. doi:https://doi.org/10.1016/j.cognition.2018.10.004

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4.* Retrieved from http://arxiv.org/abs/1406.5823

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141. doi:https://doi.org/10.1016/J.NEURON.2005.05.020

Berggren, M., Akrami, N., Bergh, R., & Ekehammar, B. (2019). Motivated social cognition and authoritarianism: Is it all about closed-mindedness? *Journal of Individual Differences* doi:https://doi.org/10.1027/1614-0001/a000293

Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, *45*(4), 792–804. doi:https://doi.org/10.1037/0022-3514.45.4.792

Butler, A. C., Fazio, L. K., & Marsh, E. J. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review*, *18*(6), 1238–1244. doi:https://doi.org/10.3758/s13423-011-0173-y

Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, *105*(2), 290–298. doi:https://doi.org/10.1037/a0031026

Butler, J. C. (2000). Personality and emotional correlates of right-wing authoritarianism. *Social Behavior and Personality*, *28*(1), 1–14. doi:https://doi.org/10.2224/sbp.2000.28.1.1

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1491–1494. doi:https://doi.org/10.1037/0278-7393.27.6.1491

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*(1), 417–440. doi:https://doi.org/10.1146/annurev.ps.41.020190.002221

Duckitt, J. (2001). A dual-process cognitive-motivational theory of ideology and prejudice. *Advances in Experimental Social Psychology*, *33*, 41–113. doi:https://doi.org/10.1016/s0065-2601(01)80004-6

Duckitt, J., & Bizumic, B. (2013). Multidimensionality of right-wing authoritarian attitudes: Authoritarianism-conservatism-traditionalism. *Political Psychology*, *34*(6), 841–862. doi:https://doi.org/10.1111/pops.12022

Duckitt, J., Bizumic, B., Krauss, S. W., & Heled, E. (2010). A tripartite approach to right-wing authoritarianism: The authoritarianism-conservatism-traditionalism model. *Political Psychology*, *31*(5), 685–715. doi:https://doi.org/10.1111/j.1467-9221.2010.00781.x

Ecker, U. K. H., & Ang, L. C. (2019). Political Attitudes and the Processing of Misinformation Corrections. *Political Psychology*, *40*(2), 241–260. doi:https://doi.org/10.1111/pops.12494

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

Ecker, U. K. H., Hogan, J. L., & Lewandowsky, S. (2017). Reminders and repetition of misinformation: Helping or hindering its retraction? *Journal of Applied Research in Memory and Cognition*, *6*(2), 185–192. doi:https://doi.org/10.1016/j.jarmac.2017.01.014

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*(1), 20–23. doi:https://doi.org/10.1177/0963721410396620

Häkkinen, K., & Akrami, N. (2014). Ideology and climate change denial. *Personality and Individual Differences*, *70*, 62–65. doi:https://doi.org/10.1016/j.paid.2014.06.030

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*(3), 188–201.

Harbour, K. E., Evanovich, L. L., Sweigart, C. A., & Hughes, L. E. (2015). A brief review of effective teaching practices that maximize student engagement. *Preventing School Failure*, *59*(1), 5–13. doi:https://doi.org/10.1080/1045988X.2014.919136

Harnish, R. J., Bridges, K. R., & Gump, J. T. (2018). Predicting economic, social, and foreign policy conservatism: The role of right-wing authoritarianism, social dominance orientation, moral foundations orientation, and religious fundamentalism. *Current Psychology*, *37*(3), 668–679. doi:https://doi.org/10.1007/s12144-016-9552-x

Ho, A. K., Sidanius, J., Kteily, N., Sheehy-Skeffington, J., Pratto, F., Henkel, K. E., … Stewart, A. L. (2015). The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO7 Scale. *Journal of Personality and Social Psychology*, *109*(6), 1003–1028. doi:https://doi.org/10.1037/pspi0000033

Hodson, G., & Sorrentino, R. M. (1999). Uncertainty orientation and the Big Five personality structure. *Journal of Research in Personality*, *33*(2), 253–261. doi:https://doi.org/10.1006/jrpe.1999.2244

Hotchin, V., & West, K. (2018). Openness and Intellect differentially predict right-wing authoritarianism. *Personality and Individual Differences*, *124*, 117–123. doi:https://doi.org/10.1016/j.paid.2017.11.048

Johnson, H., & Seifert, C. (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.

Kraft, P. W., Lodge, M., & Taber, C. S. (2015). Why people "don't trust the evidence". *The ANNALS of the American Academy of Political and Social Science*, *658*(1), 121–133. doi:https://doi.org/10.1177/0002716214554758

Lauriola, M., Foschi, R., & Marchegiani, L. (2015). Integrating values and cognitive style in a model of right-wing radicalism. *Personality and Individual Differences*, *75*, 147–153. doi:https://doi.org/10.1016/j.paid.2014.11.028

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

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361–366. doi:https://doi.org/10.1101/lm.94705

Meloen, J. D. (2019). A critical analysis of forty years of authoritarianism research: Did theory testing suffer from Cold War attitudes? In R. F. Farnen (Ed.), *Nationalism, ethnicity, and identity* (pp. 127–166). New York, NY: Routledge. doi:https://doi.org/10.4324/9781315125091-4

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*(1), 465–489. doi:https://doi.org/10.1146/annurev-psych-010416-044022

Meyer, A., Frederick, S., Burnham, T. C., Guevara Pinto, J. D., Boyer, T. W., Ball, L. J., … Schuldt, J. P. (2015). Disfluent fonts don't help people solve math problems. *Journal of Experimental Psychology: General*, *144*(2), e16–e30. doi:https://doi.org/10.1037/xge0000049

Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain. *The ANNALS of the American Academy of Political and Social Science*, *658*(1), 36–66. doi:https://doi.org/10.1177/0002716214555474

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

Nyhan, B., Reifler, J., Richey, S., & Freed, G. L. (2014). Effective messages in vaccine promotion: A randomized trial. *Pediatrics*, *133*(4), e835–e842. doi:https://doi.org/10.1542/peds.2013-2365

Onraet, E., Van Hiel, A., Dhont, K., Hodson, G., Schittekatte, M., & De Pauw, S. (2015). The association of cognitive ability with right-wing ideological attitudes and prejudice: A meta-analytic review. *European Journal of Personality*, *29*(6), 599–621. doi:https://doi.org/10.1002/per.2027

Peterson, B. E., Smirles, K. A., & Wentworth, P. A. (1997). Generativity and authoritarianism: Implications for personality, political involvement, and parenting. *Journal of Personality and Social Psychology*, *72*(5), 1202–1216. doi:https://doi.org/10.1037/0022-3514.72.5.1202

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, *9*(1673), 1–14. doi:https://doi.org/10.1038/s41467-018-03992-5

Pluviano, S., Watt, C., & Della Sala, S. (2017). Misinformation lingers in memory: Failure of three pro-vaccination strategies. *PLOS ONE*, *12*(7), e0181640. doi:https://doi.org/10.1371/journal.pone.0181640

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. doi:https://doi.org/10.1037/0022-3514.67.4.741

Rescorla, R. A., & Wagner, A. R. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts. Retrieved from https://pdfs.semanticscholar.org/afaf/65883ff75cc19926f61f181a687927789ad1.pdf

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology*, *28*(24), 4014–4021. doi:https://doi.org/10.1016/j.cub.2018.10.053

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 1–8. doi:https://doi.org/10.1073/PNAS.1805871115

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.)*, *275*(5306), 1593–1599. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9054347

Southwell, B. G., & Thorson, E. A. (2015). The prevalence, consequence, and remedy of misinformation in mass media systems. *Journal of Communication*, *65*(4), 589–595. doi:https://doi.org/10.1111/jcom.12168

Stenhouse, N., Myers, T. A., Vraga, E. K., Kotcher, J. E., Beall, L., & Maibach, E. W. (2018). The potential role of actively open-minded thinking in preventing motivated reasoning about controversial science. *Journal of Environmental Psychology*, *57*, 17–24. doi:https://doi.org/10.1016/j.jenvp.2018.06.001

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (2nd). Cambridge, MA: MIT Press.

Thorson, E. (2016). Belief echoes: The persistent effects of corrected misinformation. *Political Communication*, *33*(3), 460–480. doi: https://doi.org/10.1080/10584609.2015.1102187

Van Hiel, A., Cornelis, I., & Roets, A. (2007). The intervening role of social worldviews in the relationship between the five-factor model of personality and social attitudes. *European Journal of Personality*, *21*(2), 131–148. doi:https://doi.org/10.1002/per.618

Voss, J. F., Perkins, D. N., & Segal, J. W. (Eds.). (1991). *Informal reasoning and education*. New York, NY: Routledge.

Washburn, A. N., & Skitka, L. J. (2018). Science denial across the political divide: Liberals and conservatives are similarly motivated to deny attitude-inconsistent science. *Social Psychological and Personality Science*, *9*(8), 972–980. doi:https://doi.org/10.1177/1948550617731500

Watabe-Uchida, M., Eshel, N., & Uchida, N. (2017). Neural circuitry of reward prediction error. *Annual Review of Neuroscience*, *40*(1), 373–394. doi:https://doi.org/10.1146/annurev-neuro-072116-031109

Wickham, H., & Winston, C. (2019). Create elegant data visualisations using the grammar of graphics (R Package ggplot2, Version 3.2.1([Computer software]. doi:https://doi.org/10.1093/bioinformatics/btr406

Zakrisson, I. (2005). Construction of a short version of the Right-Wing Authoritarianism (RWA) scale. *Personality and Individual Differences, 39*(5), 863–872. doi:https://doi.org/10.1016/j.paid.2005.02.026