Received: 14 December 2021 | Revised: 9 June 2022 | Accepted: 1 August 2022

DOI: 10.1002/acp.3991

RESEARCH ARTICLE

WILEY

Asymmetry in belief revision

Brenda W. Yang[1] | **Alexandria R. Stone**[2] | **Elizabeth J. Marsh**[2]

[1]Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, USA

[2]Department of Psychology and Neuroscience, Duke University, Durham, North Carolina, USA

Correspondence
Alexandria R. Stone, Department of Psychology and Neuroscience, Duke University, Durham, NC, USA.
Email: alexandria.stone@duke.edu

Abstract

Information can change: science advances, newspapers retract claims, and reccomendations shift. Successfully navigating the world requires updating and changing beliefs, a process that is sensitive to a person's motivation to change their beliefs as well as the credibility of the source providing the new information. Here, we report three studies that consistently identify an additional factor influencing belief revision. Specifically, we document an asymmetry in belief revision: people are better able to believe in a claim once thought to be false, as opposed to unbelieving something once believed to be true. We discuss how this finding integrates and extends prior research on social and cognitive contributions to belief revisions. This work has implications for understanding the widespread prevalence and persistence of false beliefs in contemporary societies.

KEYWORDS
belief revision, feedback, misinformation, negation, truth

1 | INTRODUCTION

Successfully navigating the world requires the ability to update one's beliefs. For instance, recommendations shift, such as whether aspirin should be taken daily. Federal agencies issue new nutrition guidelines every 5 years, often with drastic changes. And newspapers retract information, such as the claim that weapons of mass destruction (WMDs) were discovered during the Iraq War. Revising beliefs can be difficult, as it requires people to overcome biases in reasoning, accept new information as true, and remember those corrections over time (Butler et al., 2011; Lewandowsky et al., 2012). Here, we briefly survey past work on belief revision before focusing on another factor affecting belief revision: whether the correction aims to instill a belief in something once thought to be false or to rescind an existing belief (e.g., to *stop* believing something).

Prior work has identified several ways that social and motivational factors influence the likelihood of belief revision when new, corrective information is encountered. For example, seminal research from Hovland and Weiss (1951) found that people are more likely to discount information from low credibility sources (e.g., a gossip columnist) compared to high credibility ones (e.g., *New England Journal of Biology and Medicine*). Generally, people are more persuaded by sources that are perceived as trustworthy than sources perceived as untrustworthy (Eagly et al., 1978; Mills & Jellison, 1967; Pilditch et al., 2020; Priester & Petty, 1995; Swire & Ecker, 2018), and source trustworthiness influences the likelihood of belief revision (Guillory & Geraci, 2013; Pluviano et al., 2020; Wolf et al., 2012). Furthermore, prior beliefs and worldviews can play a powerful role in how new information is evaluated and incorporated into the knowledge base: People may be motivated to maintain certain social, political, and moral beliefs, even in the face of compelling countervailing arguments and evidence (Haidt, 2012; Kunda, 1990; Stanley, Henne, et al., 2020; Stanley, Marsh, & Kay, 2020; Taber et al., 2009; but see Swire-Thompson et al., 2020 for an alternate view).

To return to the previous example about WMDs, non-Americans (specifically Germans and Australians) were more likely to revise their beliefs about the existence of WMDs in Iraq than were Americans, and many Americans continued to believe they existed despite remembering the retractions. This cross-national difference is likely attributable to differences in the degree to which participants were skeptical of the underlying motives for beginning the war, the trustworthiness of the sources issuing the retractions, and the messaging of political elites and news media in each respective country (Lewandowsky et al., 2005, 2012).

Cognitive factors also play a role in belief revision. For example, people take shortcuts when judging whether something is true, treating ease of processing (fluency) as evidence that something is true

(Alter & Oppenheimer, 2009; Brashier & Marsh, 2020). The result is that simply repeating a statement can make it seem truer (*illusory truth*; Hasher et al., 1977), as does presenting a statement in a high contrast easy-to-read color (Park & Schwartz, 1999). Initially researchers were concerned that it would be problematic to repeat a falsehood during correction (e.g., "It is a myth that *X*"), given that doing so likely eases processing of the falsehood (Ecker et al., 2017; Lewandowsky et al., 2012; Pennycook et al., 2018). However, more recent work suggests that repetition of a falsehood can be helpful, if it is done once, clearly labeled with a warning, and presented in conjunction with an explanation of why the information is incorrect (see meta-analysis by Ecker et al., 2022).

The effectiveness of a warning paired with an explanation is consistent with the larger literature showing that the most effective corrections do more than negate incorrect information: They also offer alternative, concrete answers or explanations. This is true whether one has made an error translating a foreign-language vocabulary word (Pashler et al., 2005), remembering a specific episodic event (Ecker et al., 2015; Johnson & Seifert, 1994; Mullet & Marsh, 2016; Tenney et al., 2009), or evaluating whether there were WMD sites in Iraq (Lewandowsky et al., 2012). In all of these cases, a simple negation creates a hole in one's existing mental representation. With time, the new association (the tag that something is wrong) will be forgotten at a faster rate than the old association (Jost's Law)—meaning that the re-emergence of the error is likely over time if nothing is provided to fill that gap (Butler et al., 2011; Ecker et al., 2022).

Of course, sometimes it is not possible to provide an alternate explanation. For instance, while the scientific consensus is that vaccines do *not* cause autism (DeStefano et al., 2013; Taylor et al., 2014), there is no simple explanation available to describe the complex genetic and environmental interactions that *do* (Marsh et al., 2016). Here we examine these types of corrections, where participants are only told that a current belief is incorrect. We compare affirmations and negations, and ask whether one is easier to update than the other. By affirmation, we mean the belief that X is true, whereas a negation is the belief that X is *not* true. We test the prediction that it will be easier to start believing in X after a correction (switching from a negation to an affirmation) than to unbelieve X (switching from an affirmation to a negation).

Our predictions rest on several related cognitive literatures. First, the initial memory may be stronger in the affirmation case, as people are biased to initially believe information and tagging something as false requires a second step that requires cognitive resources (Gilbert et al., 1993). Second, negative corrections (which tell people to unbelieve something) provide neither explanations for why beliefs are false, nor new positive claims with which to replace them (Lewandowsky et al., 2012). Believing something once thought to be false should act similar to an explanation, in that it gives people something concrete to believe in. Finally, a negative correction may functionally act like a forgot cue (Mayo et al., 2014). That is, in addition to being less memorable than affirmations, negations may afford inhibitory processing.

This possible asymmetry in belief revision has mostly been investigated in the domain of event memory, using what is known as the *continued influence paradigm* (Johnson & Seifert, 1994). In a typical study, participants read a series of news dispatches (i.e., about discovering paint cans on the site of a warehouse fire) but later some of the information is retracted (i.e., a new dispatch comes in that indicates there were no paint cans). The paradigm simulates an unfolding news story, where initial reporting is updated as more facts are learned. The focus is on retractions and refutations, where initial beliefs are negated and/or replaced—rarely do these studies contain the conditions to directly compare the updating of affirmations to the updating of negations (rather, they compare different kinds of negations). Two exceptions come from Gordon et al. (2019), with mixed results. Briefly, a perpetrator of a crime was described as Aboriginal (affirmation) or not Aboriginal (negation), across conditions—and then this description was reversed in an update (to not aboriginal or to aboriginal, depending on the condition). In Study 1, there was no asymmetry: belief updating was similar whether participants were updating from an affirmation to a negation or vice versa. In Study 2, the materials were changed to refer the perpetrator as either "a Muslim" or "Not a Muslim"; in this study more updating occurred when participants were updating from a negation (not a Muslim) to an affirmation (is Muslim), as we predicted. In short, while interesting, these studies provide inconsistent evidence for our hypothesis. They also examined updating of a specific episodic memory (a crime), as opposed to updating of a general belief about the world, which cannot be assumed to operate in the same way (see Cantor et al., 2015).

In contrast to using an episodic memory paradigm, in three studies we examined belief revision using general knowledge statements. To control for prior beliefs, we chose statements that people were unlikely to have prior beliefs about, and used a three-part learning phase to instill new beliefs. The statements were unrelated to each other, as opposed to updating a key component of a single narrative. We used unfamiliar, relatively neutral stimuli to allow us to focus on an underlying cognitive, rather than motivational, factor influencing belief revision. Learning involved multiple exposure phases with distinct deep encoding tasks to encourage people to learn and develop the requisite beliefs needed for the correction phase. We manipulated whether corrections took the form of *negation-to-affirmation* (believing something previously thought to be false) versus the form of *affirmation-to-negation* (unbelieving something previously thought to be true). As an example, consider the situation in which an individual initially and incorrectly believes that "The capital city of Fiji is *not* Suva," but is then exposed to the correction that "The capital city of Fiji is Suva." Our prediction is that this *negation-to-affirmation* correction will be more likely to bring about stable and persistent belief change than its *affirmation-to-negation* complement, where the false belief that "The capital city of Fiji is called Apia," is rescinded with the correction "The capital city of Fiji is not called Apia."

2 | STUDY 1

2.1 | Method

2.1.1 | Participants

Eighty individuals voluntarily participated in this study via Amazon's Mechanical Turk (AMT) for monetary compensation. Participant

recruitment was restricted to individuals in the United States with a prior approval rating above 80%. Eight participants did not answer all questions or failed the self-reported attention evaluation (see below for more information), so data were analyzed with the remaining 72 individuals (M_{age} = 38.19, SD = 12.20, $range_{age}$ = [22–70], 26 females, 45 males). All participants reported being fluent English speakers. For all studies, we report all manipulations, conditions, and exclusions. A sensitivity power analysis revealed that we are powered to detect an effect size of $d_z = 0.39$ (two-tailed comparison between two dependent means, alpha = .05, power = .90; Faul et al., 2007). For this and subsequent studies, data were analyzed in each study only after all data were collected. The Duke University Campus Institutional Review Board approved procedures for this study and those that follow.

2.1.2 | Materials

We used 36 unfamiliar declarative statements from published work (Tauber et al., 2013; Wang et al., 2016). The statements came from diverse domains, including: entertainment, geography, history, science, and sports.

We tested the materials with a separate sample of participants to confirm that workers on AMT were generally unfamiliar with these items (N = 98, after excluding two participants for not answering all questions). We adapted instructions from a large-scale norming study on general knowledge statements (Tauber et al., 2013; the full instructions are available in Supplemental information S1). Briefly, participants were instructed to answer general knowledge questions and to search memory for the correct answers. Participants were also told that there would be no penalty for guessing. For example, the following question was presented to participants: *What is a poem written for a bride called?* The correct response to this question is *epithalamium*. As in Tauber et al. (2013), these responses were scored with leniency for misspellings (e.g., *epithelimium* was accepted as correct). On average, participants answered 6.9% of the questions correctly (median = 2.8%), indicating that participants were generally unfamiliar with the materials.

To adapt the materials for our experimental design, we used plausibly true and false versions of each item, as well as affirming and negating frames. Thus, each item had four versions. Table 1 shows two example items and their respective versions; the full set of materials is available at osf.io/mwgt2.

We screened these materials on AMT to ensure that truth ratings were similar for all four versions of each item. To this end, a separate sample of participants (N = 99, after excluding 1 participant for not answering all questions) judged the veracity of 36 statements on a 7-pt scale (1 = *definitely false*, 7 = *definitely true*). Half of the statements were true and half were false. Within each of those two respective sets, half were presented in affirming frames and half in negating frames. Thus, each participant judged nine true affirmations, nine false affirmations, nine negated statements that were true, and nine negated statements that were false. Each participant saw only one of the four versions of each statement, with the statement version fully counterbalanced across participants. The results suggest that ratings for true and false statements (in reality) were similar, regardless of their framing (see Table 2). Critically, regardless of the actual veracity of the statement, the average judged truth was close to the midpoint of 4 on the 7-pt scale. There were no significant differences in judged truth between any possible pair-wise comparisons between the four kinds of statements (all *ps* > .10, prior to correction for multiple comparisons). Thus, for our purposes, true and false items were functionally identical and ambiguous; we used true and false items simply so that we could ensure that all corrections would leave participants with truthful information. The learning phase was used to instill the desired beliefs, prior to the correction phase.

TABLE 1 Two stimuli examples

	Actual veracity of claim	
	True	**False**
Affirming frame	Michelangelo's statue of David is located in Florence. A poem written for a bride is an epithalamium.	Michelangelo's statue of David is located in Venice. A poem written for a bride is a canzone.
Negating frame	Michelangelo's statue of David is not located in Venice. A poem written for a bride is not a canzone.	Michelangelo's statue of David is not located in Florence. A poem written for a bride is not an epithalamium.

Note: Prior to the correction phase, each false statement appeared in an affirming or negating frame. During correction, the true version of each was presented in either an affirming or negating frame.

2.1.3 | Procedure

The experiment consisted of five phases: (1) *initial exposure phase*, (2) *second exposure phase*, (3) *practice phase*, (4) *correction phase*, and (5) *truth judgment phase* (Figure 1). We implemented this series of phases for several reasons. For one, the exposure and correction phases are necessary to study the phenomenon of interest: belief revision. Second, we used multiple exposure phases with distinct deep encoding tasks to help participants process the unfamiliar items and encourage learning, so that people would have beliefs to correct. And third, the practice phase served two purposes: It measured whether learning of affirmations and negations was similar, and it provided an additional learning opportunity through retrieval practice paired with feedback (Roediger & Butler, 2011).

All phases occurred in a single session. We instructed participants not to use any outside resources to help them with the task (and confirmed this at the end of the experiment).

The purpose of these first two phases was to implement active tasks (interest ratings and subject categorizations) to familiarize

Frame	Actual veracity	Mean judged truth	SD	95% CI
Affirmation	True	4.22	0.89	[4.04, 4.40]
Affirmation	False	4.12	0.82	[3.95, 4.28]
Negation	True	4.29	0.81	[4.13, 4.44]
Negation	False	4.20	0.81	[4.04, 4.37]

Note: $N = 99$; Participants judged the truth of each statement on a 7-pt scale.

TABLE 2 Means and variances of judged truth as a function of frame (affirmation vs. negation) and the actual veracity of the statements (true vs. false)

FIGURE 1 Study 1 experimental paradigm. Examples are shown from both *negation-to-affirmation* and *affirmation-to-negation* conditions. In the fifth phase, items were counterbalanced such that participants either saw the true or false version of the statement. (*T*) indicates the factually true version of the statement.

participants with the statements that would later be corrected, given that we designed these statements to be unfamiliar to participants. We chose the particular ratings because they are both deep processing tasks that mimic everyday judgments, as people regularly judge how interesting they find things (phase 1) and make meaning-based classifications (phase 2). In the *initial exposure phase*, participants rated each of 36 statements (18 affirmations, 18 negations) for subjective interest on a 6-point scale ranging from 1 (*very uninteresting*) to 6 (*very interesting*). To encourage the processing of the statements, participants were required to wait a minimum of 2 s before moving on to the next statement. All statements were in fact false, but participants did not receive any information regarding their veracity. We made the decision to have all initial statements be false in reality so that all corrections would, in fact, leave participants with the true beliefs. This decision was made possible by the pre-testing that showed that the true and false versions of the statements were functionally equivalent, with all average ratings close to the midpoint of the scale (see Table 2).

In the *second exposure phase*, participants saw the same 36 statements from the previous phase. To further engage subjects in semantic processing, participants were ask to classify each statement as to its subject in school: geography, history, science, English, government, or other. Participants were once again required to wait a minimum of 2 s before moving on to the next statement. As in the first phase, all statements were in fact false, but participants did not receive any information regarding their veracity.

The purpose of the *practice phase* was to boost learning of the statements as well as confirm that participants remembered similar proportions of affirmed and negated statements. Otherwise, any asymmetry found in the later phases could be ascribed to initial differences in learning, rather than the structural asymmetry we predict. Thus, in the third phase, participants identified which of two alternatives (the affirmed and negated versions) they had seen earlier. For example, if a participant saw "A poem written for a bride is not an epithalamium" (a false negation) in the first two phases, the stem for the third phase would be, "A poem written for a bride is…" Participants would then choose to fill in the blank with either "an epithalamium" or "not an epithalamium." Note that in this case, participants should choose "not an epithalamium" in order to choose the item seen in previous phases. Participants were again required to wait a minimum of 2 s before moving on to the next statement. Participants received immediate feedback after each trial indicating whether their response was correct, in the sense that they correctly remembered what they had seen in the first two phases. Note that participants' measured performance

during this third phase likely underestimates their actual subsequent memory, because they received feedback on their performance after each trial—and feedback boosts learning (Pashler et al., 2005).

After the practice phase, participants completed a 2-min mental-math distractor task.

In the fourth phase, the *correction phase*, we told participants that many of the statements shown previously were actually false. Participants were then told that they would only see true statements that would correct the false information presented earlier in the session. After reading the instructions, participants rated the corrected versions of the 36 statements for subjective interest on the same 6-point scale ranging from 1 (*very uninteresting*) to 6 (*very interesting*). Participants were required to wait at least 2 s before moving on to the next statement. Critically, in this phase, the 18 items previously presented as affirmations were shown as negations, forming the *affirmation-to-negation* condition. Conversely, items previously presented as negations were shown as affirmations, forming the *negation-to-affirmation* condition.

In the fifth and final phase, the *truth judgment phase*, participants made binary true or false judgments for all statements. The form of the statements was counterbalanced such that half of the trials represented the *affirmation-to-negation* condition, and the other half of the trials represented the *negation-to-affirmation* condition. Within each of these conditions, participants saw either the true or the false version of the statement. In other words, half of the statements that participants judged were in fact true, and the other half were false. Importantly, prior to entering this final phase of the study, participants had been explicitly told what the true version of each item was—regardless of whether it was an affirmation or negation.

At the end, participants completed a self-reported attention evaluation: "Do you feel that you paid attention, avoided distractions, and took the survey seriously?" They responded by selecting one of the following: (1) no, I was distracted; (2) no, I had trouble paying attention; (3) no, I did not take the study seriously; (4) no, something else affected my participation negatively; or (5) yes. We assured participants that their responses would not affect payment or eligibility for future studies. Only those participants who selected (5) were included in the analyses. This same self-reported attention question has been used in published research (Stanley et al., 2019; Stanley, Henne, et al., 2020; Stanley, Marsh, & Kay, 2020). Upon completion, participants were monetarily compensated for their time.

2.2 | Results

2.2.1 | Pre-correction memory for affirmed and negated statements

Were affirmed and negated statements remembered differently prior to correction? We investigated whether participants correctly remembered whether statements were affirmed or negated in the third, practice phrase. On average, memory performance was numerically higher for affirmed statements than negated ones, but this 3% difference was not significant. A paired-samples *t*-test showed no statistically significant difference in memory performance between affirmed ($M = .79$ proportion correct, SD = .19) and negated ($M = .77$ proportion correct, SD = .21) statements ($t(71) = .84$, $p = .403$, 95% CI = [−.04, .09]). Thus, we did not find evidence for the possible alternative explanation that, prior to the correction phase, participants better remember affirmed than negated statements.

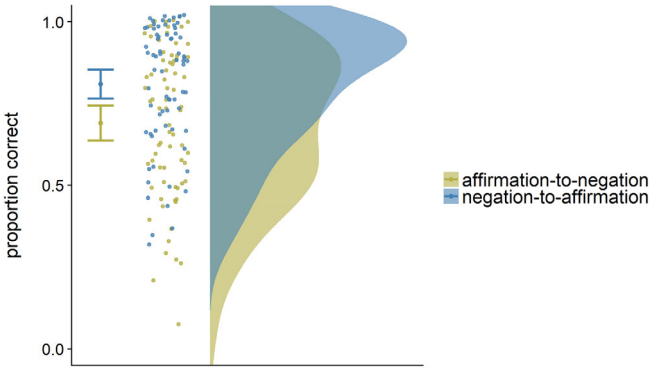

FIGURE 2 Proportion correct for affirmation-to-negation and negation-to-affirmation in Study 1. Error bars show 95% confidence intervals around the mean. Points are jittered responses of individual participants. Distributions are halved violin plots generated using ggplot2 (Wickham, 2016).

2.2.2 | Asymmetries in belief revision

Our main question was whether some beliefs are easier to revise than others, with the prediction that successful correction would be more likely to occur in the *negation-to-affirmation* direction rather than when an existing, positive belief was rescinded (the *affirmation-to-negation* direction). This prediction was corroborated: the two types of corrections led to significantly different performance on the final test, $t(71) = 4.79$, $p < .001$, 95% CI = [.07, .17], Cohen's $d_z = .56$. Performance was significantly better, on average, in the *negation-to-affirmation* condition ($M = .81$ proportion correct, SD = .19) compared to the *affirmation-to-negation* condition ($M = .69$ proportion correct, SD = .23).[1] This key finding is shown in Figure 2 as a modified raincloud plot, including 95% confidence intervals, individual participants, and density plots (Allen et al., 2019). Table 3 depicts means and variances for the proportion correct on the final truth judgment as a function of condition (*affirmation-to-negation* vs. *negation-to-affirmation*) and the correct response on the final truth judgment (true vs. false).

3 | STUDY 2

Study 1 provides initial support for our hypothesis that it is easier to believe something once thought to be false to unbelieve something. We found no positive support for the alternative hypothesis that, prior to being corrected, affirmations are more memorable than negations (as measured in phase 3 of the experiment). In Study 2,

Condition	Correct response	Mean proportion correct	SD	95% CI
Affirmation-to-negation	True	.70	.24	[.64, .76]
Affirmation-to-negation	False	.68	.27	[.61, .74]
Negation-to-affirmation	True	.85	.18	[.80, .89]
Negation-to-affirmation	False	.77	.24	[.72, .83]

Note: $N = 72$.

TABLE 3 Means and variances for proportion correct on final truth judgment as a function of condition and the correct response on the final truth judgment (true or false) in Study 1

we sought to replicate this asymmetry in belief revision and to investigate another alternative explanation for our pattern of results: that people may be more likely to believe that the repeated, affirmed statements are true (relative to repeated, negated statements) *prior* to the correction phase, making them harder to then correct.

Our pre-testing suggested minimal differences in judged truth between affirmations and negations—at least with a single exposure to the statements and with truth judgments made on a 7-pt scale. Nevertheless, previous work does suggest that negations are more complex to process (Mayo et al., 2004, 2014; Weil et al., 2020), meaning that the repeated, negated statements are likely processed with less fluency than the repeated, affirmed statements. Because fluency is a heuristic for truth (Alter & Oppenheimer, 2009; Brashier & Marsh, 2020; Fazio et al., 2015; Lewandowsky et al., 2012), the increased fluency with which the affirmations are processed may induce participants to become more likely to judge those affirmations as true, relative to the negations. To test this hypothesis in a within-subjects fashion, we made one important change to the paradigm from Study 1: The phase immediately before the correction was changed from a test measuring memory for affirmations/negations to an initial binary truth judgment rating.

3.1 | Method

3.1.1 | Participants

Eighty-two individuals voluntarily participated in this study via AMT for monetary compensation. Participant recruitment was restricted to individuals in the United States with a prior approval rating above 80%. Six participants did not answer all questions or failed the self-reported attention evaluation (detailed below); data were analyzed with the remaining 76 individuals ($M_{age} = 37.63$, $SD = 12.53$, $range_{age} = [19–72]$, 36 females, 40 males). All participants reported being fluent English speakers. We chose this sample size to match that of Study 1. A sensitivity power analysis revealed that we are powered to detect an effect size of $d_z = 0.38$ (two-tailed comparison between two dependent means, alpha = .05, power = .90).

3.1.2 | Materials

We used the same set of 36 items from Study 1.

3.1.3 | Procedure

The procedure for Study 2 was the same as Study 1, except that the third *practice phase* was replaced with an *initial truth judgment phase*. Thus, the five phases of Study 2 were the (1) *initial exposure phase*, (2) *second exposure phase*, (3) *initial truth judgment phase*, (4) *correction phase*, and (5) *final truth judgment phase*.

Following the initial exposure phase (where participants saw 36 statements and made interest ratings) and the second exposure phase (where participants indicated which subject in school they might learn about each statement), participants indicated their belief in each statement, labeling each as true or false. In this way, we investigated the judged veracity of affirmed and negated statements, respectively, prior to the correction phase. Immediately after this third phase, participants completed a 2-min mental-math distractor task.

The procedure for the remaining two phases—the correction and final truth judgment phases—were identical to Study 1. Participants also answered the same attention question described in Study 1. Upon completion, participants were monetarily compensated for their time.

3.2 | Results

3.2.1 | Initial belief in affirmed and negated statements

We first investigated whether there were differences in beliefs for affirmed versus negated statements prior to correction. A paired-samples *t*-test revealed that there was not a statistically significant difference in initial judgments about the veracity of affirmed ($M = .62$ proportion judged true, SD = .20) and negated ($M = .61$ proportion judged true, SD = .22) statements, $t(75) = .29$, $p = .774$, 95% CI = [−.06, .08]. Thus, we did not find evidence for the alternative explanation that a bias to believe affirmations drove our results. These data also provide support for our assumption that the learning phase successfully instilled beliefs, a prerequisite for correction.

3.2.2 | Asymmetries in belief revision

Next, we examined our primary effect of interest: an asymmetry in belief following corrections in the *negation-to-affirmation* direction

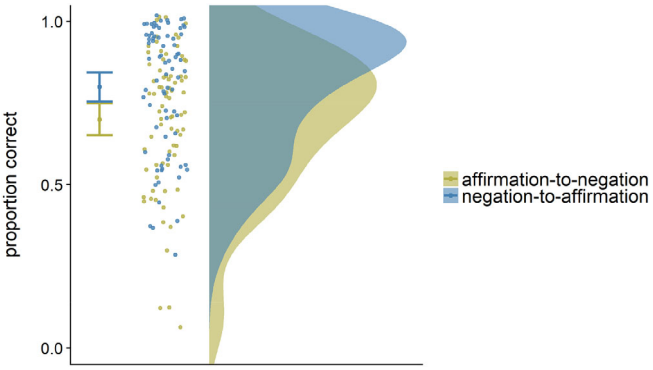

FIGURE 3 Proportion correct for affirmation-to-negation and negation-to-affirmation in Study 2. Depicted are error bars around means (95% confidence intervals), responses of individual participants, and a distribution of responses using ggplot2 (Wickham, 2016).

(believing something once thought to be false), compared to the *affirmation-to-negation* direction (rescinding belief). As in Study 1, final test performance differed significantly across the two conditions ($t(75) = 3.54$, $p = .001$, 95% CI = [.04, .16], Cohen's $d_z = .40$). We successfully replicated the finding that performance was significantly better, on average, in the *negation-to-affirmation* condition ($M = .80$ proportion correct, SD = .19) compared to the *affirmation-to-negation* condition ($M = .70$ proportion correct, SD = .21).[2]

See Figure 3 for a visual representation of these results. Table 4 depicts means and variances for the proportion correct on the final truth judgment as a function of condition (*affirmation-to-negation* vs. *negation-to-affirmation*) and the correct response on the final truth judgment (True vs. False).

4 | STUDY 3

Studies 1 and 2 corroborate the hypothesis that successful belief revision is more likely to occur in the *negation-to-affirmation* direction (when belief was instilled in something once thought to be false) than in the *affirmation-to-negation* direction (when belief was rescinded). However, because all studied statements were corrected in phase 4 of the experiment, this experimental design leaves open the possibility of heuristic responding. That is, participants could have simply endorsed the opposite of all studied statements, rather than truly engaging in a process of the belief revision (as updating is more likely when participants explicitly compare current and past information; Wahlheim & Jacoby, 2013; Wahlheim et al., 2020). While this explanation does not predict an asymmetry between conditions (with believing something once thought to be true being more successful than unbelieving something), we wanted to rule out heuristic responding as a possible explanation for our results. Study 3 addresses this potential issue by only correcting half of the initially presented statements, eliminating the possibility of heuristic responding.

4.1 | Method

4.1.1 | Participants

One hundred forty-four individuals voluntarily participated in this study via AMT for monetary compensation. Participant recruitment was restricted to individuals in the United States with a prior approval rating above 80%. Eight participants did not answer all questions or failed the self-reported attention evaluation (detailed below), so data were analyzed with the remaining 136 individuals ($M_{age} = 35.15$, SD = 9.08, $range_{age} = [21–70]$, 59 females, 76 males). All participants reported being fluent English speakers. We increased our sample size compared to previous samples, because by introducing uncorrected statements into the study design, we reduced the number of trials included in our key analyses for each participant. We balanced this change with financial constraints. A sensitivity power analysis for our primary analysis revealed that we are powered to detect an effect size of $d_z = 0.28$ (two-tailed comparison between two dependent means, alpha = .05, power = .90).

4.1.2 | Materials

We used 32 of the 36 items from Studies 1 and 2. We removed four items randomly from the 36 used in previous studies, to obtain a total number of items divisible by eight for counterbalancing. There were eight different counterbalances that served to cross (1) whether or not a given item was originally studied as an affirmation or negation; (2) whether or not it was corrected; and (3) whether the final test question was actually true or false in reality. The key manipulation is the first one: whether participants were initially exposed to an affirmation or a negation. The key outcome measure is the correction of these items (defined as accuracy on the final test, collapsing over whether the final test items were, in fact, true or false in reality). Non-corrected items are not themselves of interest but rather serve as fillers during the correction phase, so that not everything would be corrected and that heuristic responding would not be possible.

4.1.3 | Procedure

The critical difference in Study 3 relative to the previous studies was that only half of the initially presented statements were corrected in phase 4. The general procedure for Study 3 mirrored that of Study 1, except for changes in phases 3 and 4. In addition to correcting only half of the initially presented statements, we also slightly altered the structure of the third practice phrase in Study 3 to help participants encode the initial statements more deeply.

In the first phase of the experiment, the *initial exposure* phase, participants saw 32 statements. As in the previous studies, half of the statements were presented as affirmations and half as negations. Additionally, half of the initially presented statements were true and the other half were false, although participants were unaware of this

Condition	Correct response	Mean proportion correct	SD	95% CI
Affirmation-to-negation	True	.74	.22	[.69, .79]
Affirmation-to-negation	False	.66	.25	[.60, .72]
Negation-to-affirmation	True	.84	.20	[.79, .88]
Negation-to-affirmation	False	.76	.23	[.71, .81]

Note: $N = 76$.

TABLE 4 Means and variances for proportion correct on final truth judgment as a function of condition and the correct response on the final truth judgment (true or false) in Study 2

distinction (as confirmed with pre-screening of materials). We implemented this true and false division at initial exposure to allow us to correct only half of the items (those initially appearing in false frames), so as to leave participants with only true claims about the world in the correction phase, a choice made so as to try to not leave participants with false information about the world. As before, participants made interest ratings in the initial exposure phase using the same scale.

In the *second exposure phase*, participants saw all 32 statements again and indicated what subject in school they might learn about each statement. This was identical to the previous studies.

In the third phase, the *practice phase*, participants were exposed to each claim again. As in previous studies, the purpose of the practice phase was to boost learning of the statements. Thus, after being presented in each claim again, participants were randomly assigned one of two tasks on the following page. Participants were instructed to type in the last word of the statement they just saw *or* to indicate whether the word "not" was included in the statement. The specific task assigned was randomized so that participants would not be able to attend to only one part of the claim each time. Then, participants were provided with the original claim and their own responses from the previous page, and they self-scored their own responses.

In the fourth phase, the *correction phase*, we told participants that half of the statements they saw prior to the correction phase were "actually false" and the other half of the statements were "actually true"; exact, explicit instructions are available in Supporting information S1. Thus, half of the 32 initially presented statements were corrected—that is, eight affirmations and eight negations (all falsehoods)—were corrected so that they reflected the true state of the world. The remaining uncorrected statements were re-presented as in previous phases. For both corrections and re-representations, participants provided interest ratings using the same scale as in the *initial exposure* phase. Thus, this phase prevented heuristic responding, as the critical corrected items were camouflaged with filler items (ruling out the heuristic "I should just remember the opposite of what I originally studied").

In the fifth phase, the *final truth judgment phase*, participants made binary true or false judgments as in Studies 1 and 2. Accuracy was defined as proportion correct on these items, collapsing over whether the final items were true or false in reality. After completing the final truth judgment phase, participants answered the same self-reported attention evaluation question as in previous studies. Upon completion, participants were monetarily compensated for their time.

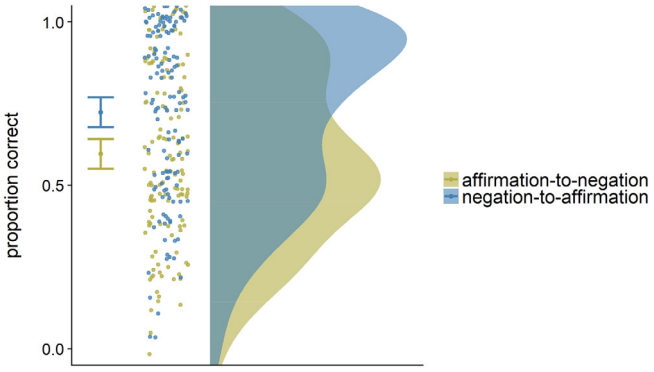

FIGURE 4 Proportion correct for affirmation-to-negation and negation-to-affirmation in Study 3. Depicted are error bars around means (95% confidence intervals), responses of individual participants, and a distribution of responses using ggplot2 (Wickham, 2016).

4.2 | Results

Consistent with our previous two studies, we found that *negation-to-affirmation* statements (i.e., that instilled belief in something once thought to be false) were more likely to be successfully corrected than *affirmation-to-negation* statements (i.e., that rescinded belief); $t(135) = 5.32$, $p < .001$, 95% CI = [.08, .18], Cohen's $d_z = .44$. Thus, we successfully replicated the finding that performance was significantly better, on average, in the *negation-to-affirmation* condition ($M = .72$ proportion correct, SD = .26) compared to the *affirmation-to-negation* condition ($M = .60$ proportion correct, SD = .26). These results are graphically depicted in Figure 4 and replicate the findings from our previous studies. For statements that were corrected, Table 5 depicts means and variances for the proportion correct in the final truth judgment phase as a function of the initial statement (affirmation vs. negation) and the correct response on the final truth judgment (true vs. false).

For completeness, we analyzed final test performance for statements that were *not* corrected (i.e., fillers). These items were processed three times in the study phase before being explicitly labeled as true for their 4th presentation, in the correction phase. On the final test, performance was higher for statements that had consistently appeared four times as affirmations, as compared to negations (see supplement for full details, Table S1, and discussion).

TABLE 5 For statements that were corrected, means and variances for proportion correct on final truth judgment as a function of the initial statement (affirmation vs. negation) and the actual correct response on the final truth judgment (true vs. false) in Study 3

Condition	Correct response on final test	Mean prop. Correct	SD	95% CI
Affirmation-to-negation	True	.67	.29	[.63, .72]
Affirmation-to-negation	False	.52	.35	[.46, .58]
Negation-to-affirmation	True	.80	.27	[.76, .85]
Negation-to-affirmation	False	.65	.36	[.58, .71]

Note: $N = 136$.

5 | GENERAL DISCUSSION

In three studies, we found evidence for an asymmetry, with people better able to correct false beliefs in the *negation-to-affirmation* direction than in the *affirmation-to-negation* direction. That is, people are better at believing something that they once thought to be false than at unbelieving something once thought to be true. It is easier to replace the belief "A poem written for a bride is *not* an epithalamium" with the belief that "A poem for a bride is an epithalamium" than to unbelieve something once thought to be true (e.g., when "Michelangelo's statue of David is located in Venice" is corrected to "Michelangelo's statue of David is *not* located in Venice").

Why are people better at correcting false beliefs in the *negation-to-affirmation* direction than in the *affirmation-to-negation* direction? In other words, why is it easier to start believing in something than to stop believing in it? Across studies, we eliminated several possible explanations. Prior to the correction phase, affirmations were not more believable (Study 2) or memorable (Study 1) than negations. Study 3 suggests that our results are not attributable to a decision-making heuristic that participants could have used in Studies 1 and 2—namely, that participants might have endorsed the opposite of whatever was initially presented (a possibility when all studied items were corrected).

Because prior work shows that some negations are more useful than others (Mayo et al., 2004), we conducted a brief follow-up in which we assessed the perceived usefulness of our statements. A separate group of participants ($N = 60$, no exclusions) was recruited through Amazon's Mechanical Turk. They were instructed to imagine future situations in which it would be useful to know the various affirmed and negated statements used in studies described here. Each participant saw 18 affirmed and 18 negated statements, all of which were true in reality. Participants rated the usefulness of knowing each statement on a scale from 1 (*definitely not useful*) to 6 (*definitely useful*). We found that participants judged affirmations ($M = 4.01$, $SD = .98$) to be more useful, on average, than negations ($M = 2.86$, $SD = 1.29$; $t(59) = 7.43$, $p < .001$, 95% CI = [.84, 1.45]). At a cognitive level, the perceived usefulness of information likely translates into increased attention to it, which in turn promotes two things: remembering and the likelihood of noticing that the information has changed. The latter would be consistent with past work showing that corrections are more likely when a participant engages in a comparison process and notices the difference between incoming information and one's original belief (remindings; Braasch & Bråten, 2017; Jacoby et al., 2015; Pashler et al., 2013).

Beyond attentional effects at encoding, these types of negations leave holes in the resulting mental model. Consider a learner who believes that the statue of David is located in Venice but learns that this belief is incorrect. A simple negation (not Venice) leaves open the possibility that the statue could be in Rome, Naples, Verona, Milan, or any other of a multitude of cities. Such a negation would be classified as *uni-polar* or *un-bounded* (Gordon et al., 2019) because it does not imply a specific alternate state. In contrast, the learner who initially believed that the statue is *not* in Florence started with a mental model with missing location information—and learning that the statue is in fact in Florence fills the gap and yields a coherent mental model. In addition to being more memorable (citation), coherence itself is interpreted as evidence for that something is true (Unkelbach & Rom, 2017).

We documented an asymmetry using relatively weak beliefs; it has not been directly tested in studies targeting major misconceptions. Perhaps the most similar study comes from Horne et al. (2015), who targeted false beliefs and negative attitudes about vaccination, an important public health misconception that is notoriously difficult to correct. Participants in this study received information about the dangers of diseases like measles and mumps (i.e., instilling a new, affirmative belief about disease risk) rather than messages debunking their belief in the MMR vaccine—autism link (i.e., belief negation). As we would predict, Horne et al. (2015) had more success correcting anti-vaccine attitudes when they attempted to instill a new, positive belief (disease risk) than when they attempted to rescind a pre-existing belief (the vaccine-autism link). This particular intervention supports our claim of an asymmetry in belief revision, although it was not clear how much participants knew about the side effects of measles and mumps, making it hard to disentangle belief revision from the effects of learning new information.

Relatedly, future research may investigate whether the confidence with which participants hold their priors affects the likelihood of belief revision in our proposed framework. The statements used in our studies were selected because very few people actually know—based on extensive pre-testing—whether they are true upon initial exposure. Our intention was to create a similar starting point for all participants, and then to repeat the statements several times to encourage participants to believe that they were true before correction (by taking advantage of the illusory truth effect; Dechêne et al., 2010; Hasher et al., 1977). However, not all participants believed the repeated statements were true before correction, likely adding noise to our data and making it harder to find our hypothesized effects. Starting instead with common misconceptions confidently held by much of public (e.g., that sugar causes hyperactivity in

children) could increase both ecological validity and the size of our effects.

In summary, we provide empirical evidence for an asymmetry in belief revision, documenting an underlying structural principle that predicts the likelihood of belief revision. We see the present research as contributing to an existing body of work describing a constellation of cognitive factors relevant to the correction of false information, such as story coherence, fluency, source credibility, and more (Brashier & Marsh, 2020; Lewandowsky et al., 2012; Lorenz-Spreen et al., 2020). Future work should aim to move from the necessary initial step of empirically describing such effects to pulling these pieces into a larger theoretical framework of belief revision. Such theory-making will require knitting these documented cognitive effects with existing frameworks from social psychology (e.g., the Elaboration Likelihood Model of persuasion; Petty & Cacioppo, 1986). However, models of persuasions such as the ELM do not include an asymmetry in correcting affirmations and negations as a factor. As we move to combining cognitive, social, and other literatures for a fuller understanding of how people's beliefs change, we advocate for the inclusion of this factor in belief revision.

ACKNOWLEDGMENTS

This work was supported by a collaborative award from the James S. McDonnell Foundation to EJM. We thank Camila Vargas Restrepo for assistance with data visualizations and members of the Marsh Lab for helpful comments on earlier versions of this manuscript. We also want to thank Matthew L. Stanley for his involvement on the project.

CONFLICT OF INTEREST

The authors declare that they have no competing interests with respect to the publication of this article.

DATA AVAILABILITY STATEMENT

De-identified data are publicly available on OSF: osf.io/mwgt2. None of these studies were formally pre-registered.

ORCID

Alexandria R. Stone https://orcid.org/0000-0002-2297-6755

ENDNOTES

[1] Chance performance is 50% correct. In both conditions (*affirmation-to-negation* and *negation-to-affirmation*), performance was above chance, on average (*ps* < .05).

[2] Note that chance performance is 50% correct. In both conditions (*affirmation-to-negation* and *negation-to-affirmation*), performance was above chance, on average (*ps* < .05).

REFERENCES

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Welcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

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. https://doi.org/10.1177/1088868309341564

Braasch, J. L., & Bråten, I. (2017). The discrepancy-induced source comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist*, *52*(3), 167–181.

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*, 499–515.

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. https://doi.org/10.3758/s13423-011-0173-y

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition*, *43*(2), 193–205.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257.

DeStefano, F., Price, C. S., & Weintraub, E. S. (2013). Increasing exposure to antibody-stimulating proteins and polysaccharides in vaccines is not associated with risk of autism. *The Journal of Pediatrics*, *163*(2), 561–567. https://doi.org/10.1016/j.jpeds.2013.02.001

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424–435. https://doi.org/10.1037/0022-3514.36.4.424

Ecker, U. K., 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.

Ecker, U. K., Lewandowsky, S., Cheung, C. S., & Maybery, M. T. (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. *Journal of Memory and Language*, *85*, 101–115.

Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Panayiota, K., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, *1*(1), 13–29.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

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

Gilbert, D. T., Tafarodi, R. W., & Malone, P. S. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, *65*(2), 221–233.

Gordon, A., Ecker, U. K. H., & Lewandowsky, S. (2019). Polarity and attitude effects in the continued-influence paradigm. *Journal of Memory and Language*, *108*, 104028.

Guillory, J. J., & Geraci, L. (2013). Correcting erroneous inferences in memory: The role of source credibility. *Journal of Applied Research in Memory and Cognition*, *2*(4), 201–209.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, *16*, 107–112.

Horne, Z., Powell, D., Hummel, J. E., & Holyoak, K. J. (2015). Countering antivaccination attitudes. *Proceedings of the National Academy of Sciences*, *112*(33), 10321–10324. https://doi.org/10.1073/pnas.1504019112

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, *15*(4), 635–650.

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. M. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(5), 1282–1297.

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.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

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. https://doi.org/10.1177/1529100612451018

Lewandowsky, S., Stritzke, W. G. K., Oberauer, K., & Morales, M. (2005). Memory for fact, fiction, and misinformation: The Iraq war 2003. *Psychological Science*, *16*(3), 190–195. https://doi.org/10.1111/j.0956-7976.2005.00802.x

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, *4*(11), 1102–1109.

Marsh, E. J., Cantor, A. D., & Brashier, M. (2016). Believing that humans swallow spiders in their sleep. In *Psychology of learning and motivation* (Vol. 64, pp. 93–132). Academic Press. http://linkinghub.elsevier.com/retrieve/pii/S0079742115000341

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*(4), 433–449. https://doi.org/10.1016/j.jesp.2003.07.008

Mayo, R., Schul, Y., & Rosenthal, M. (2014). If you negate, you may forget: Negated repetitions impair memory compared with affirmative repetitions. *Journal of Experimental Psychology: General*, *143*(4), 1541–1552. https://doi.org/10.1037/a0036122

Mills, J., & Jellison, J. M. (1967). Effect on opinion change of how desirable the communication is to the audience the communicator addressed. *Journal of Personality and Social Psychology*, *6*(1), 98–101. https://doi.org/10.1037/h0021217

Mullet, H. G., & Marsh, E. J. (2016). Correcting false memories: Errors must be noticed and replaced. *Memory & Cognition*, *44*(3), 403–412. https://doi.org/10.3758/s13421-015-0571-x

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8.

Pashler, H., Kang, S. H., & Mozer, M. C. (2013). Reviewing erroneous information facilitates memory updating. *Cognition*, *128*(3), 424–430.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, *147*(12), 1865–1880.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Pilditch, T. D., Madsen, J. K., & Custers, R. (2020). False prophets and Cassandra's curse. The role of credibility in belief updating. *Acta Psychologica*, *202*, 102956.

Pluviano, S., Della Sala, S., & Watt, C. (2020). The effects of source expertise and trustworthiness on recollection: The case of vaccine misinformation. *Cognitive Processing*, *21*(3), 321–330.

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

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Stanley, M. L., Henne, P., Yang, B. W., & De Brigard, F. (2020). Resistance to position change, motivated reasoning, and polarization. *Political Behavior*, *42*(3), 891–913.

Stanley, M. L., Marsh, E. J., & Kay, A. C. (2020). Structure-seeking as a psychological antecedent of beliefs about morality. *Journal of Experimental Psychology: General*, *149*(10), 1908–1918.

Stanley, M. L., Yin, S., & Sinnott-Armstrong, W. (2019). A reason-based explanation for moral dumbfounding. *Judgment and Decision making*, *14*, 120–129.

Swire, B., & Ecker, U. K. (2018). Misinformation and its correction: Cognitive mechanisms and recommendations for mass communication. *Misinformation and Mass Audiences*, 195–3111.

Swire-Thompson, B., DeGutis, J., & Lazer, D. (2020). Searching for the backfires effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition*, *9*(3), 286–299.

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155. https://doi.org/10.1007/s11109-008-9075-8

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods*, *45*(4), 1115–1143.

Taylor, L. E., Swerdfeger, A. L., & Eslick, G. D. (2014). Vaccines are not associated with autism: An evidence-based meta-analysis of case-control and cohort studies. *Vaccine*, *32*(29), 3623–3629. https://doi.org/10.1016/j.vaccine.2014.04.085

Tenney, E. R., Cleary, H. M. D., & Spellman, B. A. (2009). Unpacking the doubt in "beyond a reasonable doubt": Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology*, *31*(1), 1–8. https://doi.org/10.1080/01973530802659687

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition*, *160*, 110–126.

Wahlheim, C. N., Alexander, T. R., & Peske, C. D. (2020). Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term. *Psychological Science*, *31*(10), 1325–1339.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, *41*(1), 1–15.

Wang, W.-C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On known unknowns: Fluency and the neural mechanisms of illusory truth. *Journal of Cognitive Neuroscience*, *28*(5), 739–746. https://doi.org/10.1162/jocn_a_00923

Weil, R., Schul, Y., & Mayo, R. (2020). Correction of evident falsehood requires explicit negation. *Journal of Experimental Psychology: General*, *149*(2), 290–310.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org.

Wolf, A. G., Rieger, S., & Knauff, M. (2012). The effects of source trustworthiness and inference type on human belief revision. *Thinking & Reasoning*, *18*(4), 417–440. https://doi.org/10.1080/13546783.2012.677757

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yang, B. W., Stone, A. R., & Marsh, E. J. (2022). Asymmetry in belief revision. *Applied Cognitive Psychology*, 1–11. https://doi.org/10.1002/acp.3991