RESEARCH ARTICLE

WILEY

Pictures and repeated exposure increase perceived accuracy of news headlines

Thomas J. Smelter | Dustin P. Calvillo

Psychology Department, California State University San Marcos, San Marcos, California

Correspondence
Dustin P. Calvillo, Psychology Department, California State University San Marcos, 333 South Twin Oaks Valley Road, San Marcos, CA 92096.
Email: dcalvill@csusm.edu

Summary

Fake news, deliberately inaccurate and often biased information that is presented as accurate reporting, is perceived as a serious threat. Recent research on fake news has documented a high general susceptibility to the phenomenon and has focused on investigating potential explanatory factors. The present study examined how features of news headlines affected their perceived accuracy. Across four experiments (total $N = 659$), we examined the effects of pictures, perceptual clarity, and repeated exposure on the perceived accuracy of news headlines. In all experiments, participants received a set of true and false news headlines and rated their accuracy. The presence of pictures and repeated exposure increased perceived accuracy, whereas manipulations of perceptual clarity did not show the predicted effects. The effects of pictures and repeated exposure were similar for true and false headlines. These results demonstrate that accompanying pictures and repeated exposure can affect evaluations of truth of news headlines.

KEYWORDS
cognitive reflection, fake news, illusory truth, judgment and decision making, repeated exposure

1 | INTRODUCTION

Performing a GoogleTrends search for the term *fake news* from 2004 to 2019 reveals a flat and infrequent use of the term until October 2016, the month before the 2016 US presidential election, when a drastic increase in popularity of the phrase occurred. Fake news contains fabricated content that is intended to mimic factual news (Lazer et al., 2018; Pennycook, Cannon, & Rand, 2018). Since 2016, fake news has had no shortage of media coverage, however, is there empirical evidence to suggest it is a problem?

According to a 2019 Pew Research Study, American people seem to view fake news as a serious problem. In this study, respondents reported that they viewed fake news as more of a threat than climate change, violent crime, racism, illegal immigration, terrorism, and sexism (Mitchell, Gottfried, Stocking, Walker, & Fedeli, 2019). These heightened fears may lead to accelerated polarization and may result in an increase in cynical attitudes towards news. In their responses to the statement, "The issue of made-up news and information has led

you to...," a majority of respondents reported that the situation has compelled them to completely cease accessing some news outlets and nearly half of them reported consuming less information overall. The latter finding is especially problematic as an increasingly complex world requires *more* critical consumption of information not less.

While the aforementioned finding is alarming, other recent events demonstrate how the fake news phenomenon is leading to tangible consequences like threats of physical violence and even a shooting via an AR-15. On December 4, 2016, a 28-year-old man named Edgar Welch, from Salisbury, North Carolina, stormed into a pizzeria, pointed an AR-15 at an employee and discharged it multiple times in the restaurant with none leading to injury (Hsu, 2017; Kang & Goldman, 2016). His impetus was an article he read online stating Hillary Clinton was running a child sex-slave ring out of the restaurant – Comet Ping Pong. The article, which appeared online in November 2016, right before the 2016 U.S. presidential election, was quickly exposed as false by investigations by the New York Times and the fact-checking website Snopes.com (Kang, 2016; LaCapria, 2016). In

that same period of time, shortly before the election, James Alefantis, the owner of the pizza restaurant, received numerous messages, like "we're on to you" and "I will kill you personally" (Kang, 2016).

Social media (e.g., Facebook and Twitter) seems to have been particularly influential in the acceleration of the pace at which the fake articles regarding the pizzeria spread online (Kang & Goldman, 2016). Furthermore, empirical investigations have shown social media entities to be "breeding grounds" for fake news. For example, on Facebook before the 2016 presidential election, the 20 most popular fake news stories were more widely engaged with than the 20 most popular mainstream news stories (Silverman, Strapagiel, Shaban, & Hall, 2016). In addition, among fact-checked stories on Twitter, fake stories were 70% more likely to be tweeted than real news (Vosoughi, Roy, & Aral, 2018). This proliferation of fake news presents a serious dilemma for democracies, whose success is dependent on the ability of collective, and often opposing, groups to reach enough agreement about reality in order to solve complex problems.

A few studies have examined susceptibility to fake news by providing participants with real and fake news headlines. Individuals low in analytic thinking (Pennycook & Rand, 2019), high in dogmatism and religious fundamentalism (Bronstein, Pennycook, Bear, Rand, & Cannon, 2019), and high in bullshit receptivity and high in overclaiming (Pennycook & Rand, 2020) are more susceptible to belief in fake news than their counterparts. Building off of Hasher, Goldstein, and Toppino's (1977) work on the illusory truth effect, Pennycook et al. (2018) recently found that prior exposure (also referred to as "repeated exposure") to news headlines increases their perceived accuracy, which was explained by the increased familiarity caused by prior exposure. In other words, the memory traces left over from the prior exposure may lead to a vague experience of familiarity, which can then be erroneously misattributed as increased confidence and belief in the claim.

In these fake news studies, participants judge the accuracy of headlines. To make these sorts of judgments, individuals search for relevant information in memory (Graesser & Hemphill, 1991). The quantity of information recollected and the ease with which it is recollected should affect accuracy judgments (Schwarz et al., 1991). However, other factors that influence ease of processing may also increase perceived accuracy. This could explain the results of Pennycook et al. (2018). Repeated exposure increases the ease of processing, which increases perceived accuracy.

In addition to repeated exposure increasing perceived accuracy of news headlines, other features of news media have also been found to lead to false memories of news events. For example, images seem to be particularly influential. In one study, Strange, Garry, Bernstein, and Lindsay (2011) had participants read true and false headlines, half of which had pictures seemingly related to the event and half that did not. Participants then indicated whether they remembered the event, knew about it, or neither. The pictures led to increased rates of remembering for the true events and, critically, for the false events that had never occurred. The results from Strange et al. (2011) suggest that pictures increase memory for fake news events. Remembering a false event, however, is not the same as perceiving a false news headline to be accurate. For example, *nonbelieved memories*, which occur when individuals have lost belief in a memory despite the

presence of vivid aspects of it, suggest there is a dissociation between memory and belief (Otgaar, Scoboria, & Mazzoni, 2014). Therefore, we sought to examine the effects of pictures on accuracy ratings of news headlines to determine if the results of Strange et al. (2011) generalize for ratings of perceived accuracy and when using the original photographs that accompany both true and false news headlines.

In four experiments, we investigated whether people can discern true from false news headlines. Because of the ubiquitous nature of visual stimuli accompanying assertions on social media and online generally, we were interested in the effect that such visual stimuli (e.g., pictures) might have on evaluations of truth. In Experiment 1, we examined how pictures affected judgments of perceived accuracy for news headlines. In Experiments 2A and 2B, we examined whether the perceptual clarity of headlines affected their perceived accuracy. In Experiment 3, we attempted to replicate the effect of prior exposure and fake news from Pennycook et al. (2018). Experiments 1 and 3 also included measures of cognitive reflection to examine if it correlates with ability to discern true from fake news, as previously reported by Bronstein et al. (2019) and Pennycook and Rand (2019). Together, this set of experiments examined how factors of news headlines affect their perceived accuracy.

2 | EXPERIMENT 1

In an ideal world, people would evaluate claims based on the logical argument proposed. However, and especially in cases with incomplete information, people often entrust their evaluations of the world to subjective feelings of truth (Abed, Fenn, & Pezdek, 2017; Cardwell, Lindsay, Förster, & Garry, 2017), or *truthiness* (Newman, Garry, Bernstein, Kantner, & Lindsay, 2012). Visual stimuli, like pictures, seem to be another factor that can lead to heightened perceptions of truthiness. For example, pictures have been shown to increase the believability of general knowledge claims (Fenn, Newman, Pezdek, & Garry, 2013; Newman et al., 2012), particularly in within-subjects designs (Newman et al., 2015). One explanation for this effect is that pictures often do not contain much useful information and, therefore, a vague relatedness that a photo provides can lead to erroneous associations of prior experience and knowledge.

Pictures have not only been shown to inflate truthiness, they also have been implicated in the creation of false memories. Doctored pictures of childhood events, when accompanied by guided imagery exercises, can create false memories in participants (Wade, Garry, Read, & Lindsay, 2002). Doctored pictures of political events also lead to false memories of those events (Frenda, Knowles, Saletan, & Loftus, 2013). True pictures can also create false memories (Lindsay, Hagen, Read, Wade, & Garry, 2004) and pictures can lead people to mistakenly believe that they read conclusions that were actually inferred from photos (Henkel, 2011).

In Experiment 1, we examined the effect of pictures on the perceived accuracy of news headlines. We gave participants a set of true and false headlines either with or without a picture and they rated the headlines' accuracy. We predicted that true headlines would be perceived as more accurate than false headlines (Hypothesis 1). Because previous studies have demonstrated that pictures inflate feelings of

FIGURE 1 Sample of true news headlines with pictures [Colour figure can be viewed at wileyonlinelibrary.com]
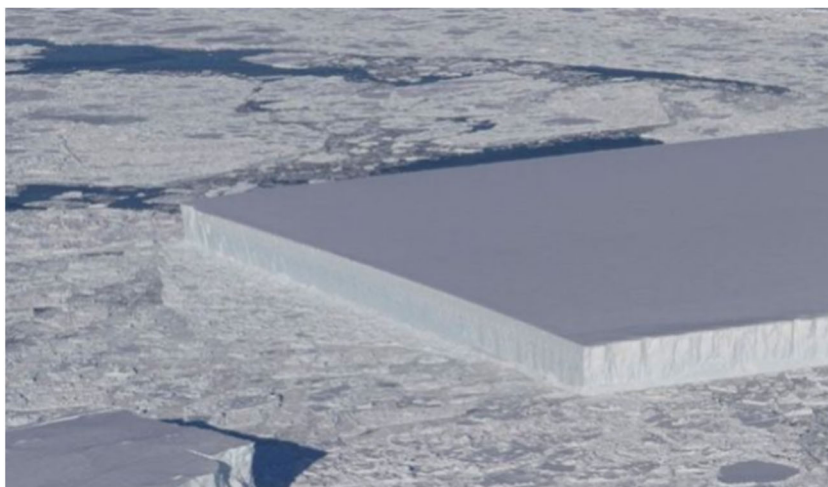
Nasa Photographs Rectangular Iceberg
Nasa has released a striking photo of a rectangular iceberg floating in the Weddell Sea off Antarctica.

Bill Gates Uses Poop as Prop to Pitch Toilet of the Future
Bill Gates shocks the audience with a jar of human excrement.

truth for general knowledge claims (e.g., Newman et al., 2012), and because we do not view knowledge discernment and news discernment tasks as different cognitive processes, we predicted that the presence of pictures would increase perceived accuracy (Hypothesis 2) for news headlines. We predicted that pictures would increase perceived accuracy for true and false headlines similarly (no interaction; Hypothesis 3). We also predicted that cognitive reflection would be positively correlated with the ability to discern true from false headlines (Hypothesis 4), replicating Pennycook and Rand (2019).

2.1 | Method

2.1.1 | Participants and design

In our preregistration, we conducted several power analyses and decided to collect data in waves until we had at least 200 participants who met inclusion criteria. A total of 289 students from a medium-sized public university in the United States completed the study online. The study included two honesty check questions (described below) and 78 participants failed at least one of the honesty check questions. Data from the remaining 211 participants (50 men, 159 women, and 2 who declined to select a gender) were included in analyses. The median age of the participants was 19 years.

This experiment used a 2 (Headline: true, false) × 2 (Picture: with, without) within-subjects factorial design. The dependent variable was perceived accuracy, which was measured on a 1 (*not at all accurate*) to 4 (*very accurate*) scale. We also measured cognitive reflection.

2.1.2 | Materials and procedure

Materials included 28 news headlines, 14 of which were false and 14 of which were true. Headlines for the false stories were sourced

Pigs Unhealthy to Eat Because They Don't Sweat

Pigs don't sweat and are therefore more likely to harbor parasites or other toxins, making them unhealthy to eat.

Monster Energy Selling Caffeinated Ham

Maybe some foods do go better with caffeine, and we'll never know until we try.

FIGURE 2 Sample of the fake news headlines with pictures [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Mean accuracy ratings (on a scale from 1 to 4) by headlines' veracity and whether it was accompanied by a picture in Experiment 1

Headline	Picture M [95% CI]	No picture M [95% CI]	Overall M [95% CI]
True	2.66 [2.59, 2.72]	2.52 [2.45, 2.58]	2.59 [2.53, 2.64]
False	2.53 [2.46, 2.60]	2.41 [2.35, 2.47]	2.47 [2.41, 2.53]
Overall	2.59 [2.54, 2.65]	2.46 [2.41, 2.52]	2.53 [2.48, 2.58]

from and had received a *false* rating by Snopes.com. True headlines were sourced from websites of mainstream news outlets (e.g., CNN, NBC News, and BBC). We selected true and false headlines that were not political in nature to reduce the confound of motivated reasoning. In regards to our selection of true news content, although we generally avoided blatantly obvious true news (e.g., who won the Superbowl), we still attempted to select the true news stimuli in a way that varied the amount of surprising and obvious headlines, so as to obtain a representative sample of headlines. For example, Figure 1 contains two of our sample true headlines: one surprising and another less surprising. The headlines and pictures were all formatted to resemble a news article link on social media. Examples are presented in Figures 1 and 2, and the entire set of headlines and source material is available in the Data S1. Half of headlines that each participant saw included a picture, while the other half did not. We counterbalanced which headlines appeared with pictures across participants.

Participants also completed a 7-item cognitive reflection test (CRT). The first four items came from Thomson and Oppenheimer (2016) and the last three came from Oldrati, Patricelli, Colombo, and Antonietti (2016). We also measured scientific reasoning, conspiracy beliefs, and active open-minding thinking. We did not, however, make any predictions about these measures. The description of these measures and their exploratory analyses, as well as the specific CRT questions that we used, are available in the Data S1.

Participants received instructions for the headline evaluation task and then rated the accuracy of 28 headlines in a random order. Then they completed the CRT, followed by the measures of active-open minded thinking, general conspiracy beliefs, and scientific reasoning. Finally, participants answered some demographic questions, two honesty check questions, and then were debriefed about the study. The honesty check questions asked participants if they responded randomly or without reading any of the questions, and if they looked up any of the headlines in the study. We preregistered the inclusion criteria that participants had to respond *no* to both of these questions to be included in the analyses.

2.2 | Results

A two-way factorial ANOVA was conducted to assess if participants were able to distinguish true from false news headlines and if pictures influenced the perceived accuracy of headlines. Mean perceived accuracy ratings for all conditions are presented in Table 1. Supporting Hypothesis 1, true headlines were perceived as more accurate than false headlines, $F(1, 210) = 23.91$, $p < .001$, $\eta_p^2 = .10$. Hypothesis 2 was also supported; The presence of pictures increased perceived accuracy, $F(1, 210) = 30.94$, $p < .001$, $\eta_p^2 = .13$. Finally, there was not a significant interaction between picture and headline conditions, consistent with Hypothesis 3, $F(1, 210) = .08$, $p = .781$, $\eta_p^2 = .00$.

To test Hypothesis 4, we computed news discernment scores by subtracting the ratings that participants gave to false headlines from the ratings that they gave to true headlines. Positive scores indicated the magnitude by which participants could discern true from false headlines. We found that news discernment was not significantly correlated with CRT performance, $r(209) = .11$, $p = .117$, failing to support Hypothesis 4. The mean number of correctly answered CRT questions was 2.76 (95% CI [2.55, 2.97]) out of 7, and the CRT had poor internal consistency (Cronbach's alpha = .53). We also performed a post hoc analysis to explore the null finding for CRT. We correlated true and false headline accuracy ratings with CRT scores. We did not find a significant correlation between CRT performance and ratings of true headlines, $r(209) = .02$, $p = .832$, or false headlines, $r(209) = −.07$,

$p = .288$. Our preregistered exploratory analyses are available in the Data S1.

3 | EXPERIMENT 2A

In Experiment 1, we found that pictures increased the perceived accuracy of news headlines. In Experiments 2A and 2B, we manipulated perceptual clarity and legibility of headlines and pictures to examine if this might affect their perceived accuracy. Perceptual clarity has been shown to affect judgments in a variety of contexts. For example, according to a study by Song and Schwarz (2008a), when instructions are written in difficult-to-read fonts, participants predict that that the task described in the instructions will take longer and they report less willingness to complete the task compared to those who receive easy-to-read instructions. Additionally, difficult-to-read fonts are posited to engage analytical thinking (Alter, Oppenheimer, Epley, & Eyre, 2007; Gervais & Norenzayan, 2012; Song & Schwarz, 2008b, but see Meyer et al., 2015). Reber & Schwarz (1999) found that presenting words in a more disfluent or difficult-to-read manner led to lower evaluations of perceived truth compared to fluent or easy-to-read font. Therefore, we wanted to test whether the manipulation of perceptual clarity of pictures and text could result in differences in perceived accuracy similar to the effect of pictures found in Experiment 1.

We had three hypotheses for Experiment 2A. We predicted that participants would rate true headlines as more accurate than false headlines (Hypothesis 1). We predicted that perceptually clear headlines would be rated as more accurate than perceptually degraded headlines (Hypothesis 2). We predicted that perceptual clarity would have a similar effect for true and false headlines (no interaction; Hypothesis 3). We considered the possibility that perceptually degraded fonts would increase analytic thinking, which would lower ratings of false headlines more than true headlines (if analytical thinking predicts news discernment). This would lead to an interaction. We decided, however, because the effect of perceptually degraded presentations on analytical thinking replicates poorly (Klein et al., 2018; Meyer et al., 2015), to predict no interaction.

3.1 | Method

3.1.1 | Participants and design

In our preregistration, we conducted a power analyses and decided to collect data until we had 120 participants. The experiment consisted of 120 undergraduate students from a medium-sized public university in the United States. There were 34 men and 86 women, and the median age of the sample was 19.5 years.

The design of Experiment 2A was a 2 (Headline: true, false) × 2 (Perceptual clarity: clear, degraded) within-subjects factorial. The dependent variable was participants' accuracy ratings on the same scale as Experiment 1.

TABLE 2 Mean accuracy ratings (on a scale from 1 to 4) by headlines' veracity and fluency in Experiments 2A and 2B

Headline	Fluent M [95% CI]	Disfluent M [95% CI]	Overall M [95% CI]
Experiment 2A			
True	2.61 [2.54, 2.67]	2.77 [2.69, 2.84]	2.69 [2.63, 2.74]
False	2.14 [2.05, 2.23]	2.29 [2.19, 2.39]	2.21 [2.14, 2.29]
Overall	2.37 [2.31, 2.44]	2.53 [2.47, 2.59]	2.45 [2.39, 2.51]
Experiment 2B			
True	3.02 [2.93, 3.10]	2.98 [2.89, 3.06]	3.00 [2.94, 3.06]
False	2.01 [1.89, 2.12]	2.12 [2.00, 2.32]	2.06 [1.98, 2.14]
Overall	2.51 [2.43, 2.59]	2.54 [2.46, 2.62]	2.52 [2.47, 2.59]

3.1.2 | Materials and procedure

A new set of 20 headlines was collected. The 10 false headlines were collected from Snopes.com, and then 10 true headlines were collected from Usnews.com. All of the participants saw all 20 headlines. Half of the headlines they saw were presented in a perceptually clear manner (five true and five false) and half were in a perceptually degraded manner (five true and five false). We counterbalanced which headlines were perceptually clear and which were perceptually degraded across participants. The perceptually clear headlines were in Myriad Pro 20-point font in black. The perceptually degraded headlines were in Myriad Pro 14-point font, in 15% gray, and in italics. These fonts were similar to those used in previous studies (e.g., Alter et al., 2007). All headlines were accompanied by a picture. The perceptually clear headlines had unaltered pictures and the perceptually degraded headlines had the same pictures but were run through a blur filter. Both versions of all headlines are included in the Data S1.

3.2 | Results

We conducted a two-way factorial ANOVA to assess the effects of headlines' truth and fluency on perceived accuracy. The mean perceived accuracy ratings for all conditions are presented in Table 2. Supporting Hypothesis 1, true headlines were perceived as more accurate than false headlines, $F(1, 119) = 185.69$, $p < .001$, $\eta_p^2 = .61$. Hypothesis 2, however, was not supported. Perceptually degraded headlines were actually perceived as more accurate than perceptually clear headlines, $F(1, 119) = 28.57$, $p < .001$, $\eta_p^2 = .19$. Finally, we did not find an interaction between veracity and perceptual clarity, $F(1, 119) = .01$, $p = .927$, $\eta_p^2 = .00$, consistent with Hypothesis 3.

4 | EXPERIMENT 2B

We found that perceptual clarity decreased perceived accuracy of news headlines in Experiment 2A, which was the opposite of our prediction. One possibility is that the within-subjects manipulation of perceptual clarity made participants aware of the purpose of the study resulting in demand characteristics. Alternatively, the awareness of the manipulation may have allowed participants to discount the effects that ease of processing may have on perceived accuracy (for similar arguments, see Oppenheimer, 2004; Topolinski & Strack, 2010). Therefore, we conducted Experiment 2B that was identical to 2A except that perceptual clarity was manipulated between-subjects.

We had the same three hypotheses as Experiment 2A: that participants would generally be able discern true from false headlines (Hypothesis 1); that headlines presented in perceptually clear font and pictures would result in greater accuracy ratings than when they were presented in perceptually degraded font and pictures (Hypothesis 2), and that there would not be an interaction (Hypothesis 3).

4.1 | Method

4.1.1 | Participants and design

In our preregistration, we conducted a power analyses and decided to collect data until we had 128 participants. A total of 142 participants from Amazon's Mechanical Turk completed this experiment. There were two honesty check questions and 14 participants failed at least one of them. The remaining 128 participants were included in our analyses. There were 71 men, 56 women, and 1 other gender, and the median age of the sample was 35.5 years.

Experiment 2B had a 2 (Headline: true, false) × 2 (Perceptual clarity: clear, degraded) mixed-model factorial design with headline within-subjects and fluency between-subjects. The dependent variable was participants' accuracy ratings on the same scale as the previous experiments.

4.1.2 | Materials and procedure

The same set of 20 headlines was used for Experiment 2B as 2A. Participants were randomly assigned into the perceptually clear or perceptually degraded condition and then rated the accuracy of all 20 headlines. The headlines were presented in a random order for each participant.

4.2 | Results

We conducted a two-way mixed-model factorial ANOVA to examine the effects of headlines' veracity and perceptual clarity on their perceived accuracy. Mean perceived accuracy ratings for all four conditions are presented in Table 2. Supporting Hypothesis 1, true headlines were perceived as more accurate than false headlines, $F(1, 126) = 422.35$, $p < .001$, $\eta_p^2 = .77$. Hypothesis 2 was not supported, perceptually clear headlines were not perceived as more accurate than perceptually degraded ones, $F(1, 126) = .30$, $p = .588$, $\eta_p^2 = .00$. Finally, we did not find a significant interaction, $F(1, 126) = 2.98$, $p = .087$, $\eta_p^2 = .02$, consistent with Hypothesis 3.

5 | EXPERIMENT 3

Experiments 2A and 2B examined manipulations related to perceptual clarity, which have been examined in other studies (Alter et al., 2007; Reber & Schwarz, 1999), in the context of news discernment. These experiments failed to find evidence that perceptual clarity increased perceived accuracy. In Experiment 3, we examined another factor that could influence accuracy ratings: repeated exposure. We had participants rate some true and false headlines for humor, complete a CRT, and then had them rate their perceived accuracy for a larger set of headlines that included those they had previously rated for humor and some they had not. Thus, we manipulated prior exposure. We

TABLE 3 Mean accuracy ratings (on a scale from 1 to 4) by headlines' veracity and whether it was previously seen in Experiment 3

Headline	Prior exposure M [95% CI]	No prior exposure M [95% CI]	Overall M [95% CI]
True	2.82 [2.77, 2.88]	2.70 [2.64, 2.76]	2.76 [2.71, 2.81]
False	2.26 [2.19, 2.33]	2.09 [2.01, 2.16]	2.17 [2.11, 2.24]
Overall	2.54 [2.49, 2.59]	2.39 [2.34, 2.45]	2.47 [2.42, 2.52]

predicted that true headlines would be perceived as more accurate than false headlines (Hypothesis 1). Based on Pennycook et al. (2018) and previous studies on the *illusory truth effect* (e.g., Fazio, Brashier, Payne, & Marsh, 2015; Hasher et al., 1977), we predicted that prior exposure would increase perceived accuracy of news headlines (Hypothesis 2). We also predicted that the effect of perceived accuracy would be similar to true and false headlines (no interaction, Hypothesis 3). In addition, we predicted that CRT scores would be related to the ability to discern true from false headlines (Hypothesis 4). Finally, we had one research question: whether people would perceive false headlines as more humorous than true headlines (Research Question 1). We compared humor ratings for false headlines to those for true headlines.

5.1 | Method

5.1.1 | Participants and design

In our preregistration, we conducted a power analyses and decided to collect data until we had 200 participants. A total of 215 participants from Amazon's Mechanical Turk completed this experiment. There were two honesty check questions and 15 participants failed at least one of them. The remaining 200 participants were included in our analyses. There were 99 men and 101 women, and the median age of the sample was 36 years.

Experiment 3 had a 2 (Headline: true, false) × 2 (Prior exposure: yes, no) within-subjects factorial design. The dependent variable was participants' accuracy ratings on the same scale as the previous experiments. We also measured humor ratings for the headlines.

5.1.2 | Materials and procedure

A set of 48 headlines included the 28 from Experiment 1 and the 20 from Experiments 2A and 2B. Half were true and half were false. Participants first rated half of the true and false headlines for humor on a 4-point scale (from *not at all funny* to *very funny*). We counterbalanced which half of the headlines were rated by participants. Next, participants completed a filler task consisting of a 9-item CRT, containing items from Toplak, West, and Stanovich (2014), Thomson and Oppenheimer (2016), Oldrati et al. (2016), Primi, Morsanyi, Chiesi, Donati, and Hamilton (2016), and Baron, Scott, Fincher, and Metz (2015). The specific CRT questions are available in

the Data S1. Finally, participants rated the accuracy of all 48 headlines, including the 24 that they had previously rated for humor and an additional 24 that they had not.

5.2 | Results

We conducted a two-way repeated measures ANOVA to examine the effects of headlines' veracity and prior exposure on their perceived accuracy. The mean perceived accuracy ratings for all conditions are presented in Table 3. Supporting Hypothesis 1, true headlines were perceived as more accurate than false headlines, $F(1, 199) = 352.73$, $p < .001$, $\eta_p^2 = .64$. Hypothesis 2 was also supported; prior exposure increased perceived accuracy, $F(1, 199) = 50.96$, $p < .001$, $\eta_p^2 = .20$. Finally, there was not a significant interaction between headline and exposure, $F(1, 199) = 1.70$, $p = .194$, $\eta_p^2 = .01$, consistent with Hypothesis 3. To test Hypothesis 4, we computed discernment scores by subtracting accuracy ratings of false headlines from those of true headlines. We then examined the correlation between discernment and CRT scores. Hypothesis 4 was also supported. There was a significant positive correlation between discernment and CRT, $r(198) = .49$, $p < .001$. The mean number of correctly answered CRT questions was 2.98 (95% CI [2.66, 3.29]) out of 9, and the CRT had acceptable internal consistency (Cronbach's alpha = .72). We performed a post hoc analysis to explore the results from CRT and discernment. We analyzed correlations for true and false headline scores with CRT separately. We did not find a significant correlation between true headline accuracy scores and CRT, $r(198) = .06$, $p = .423$. We did, however, find a significant negative correlation between false headline accuracy scores and CRT, $r(198) = −.42$, $p < .001$. This finding that the discernment correlation is influenced largely by the negative correlation with false headlines is consistent with past research (Pennycook & Rand, 2019).

To test our research question, we compared the humor ratings for true headlines to those for false headlines. False headlines ($M = 2.00$, 95% CI [1.94, 2.06]) received greater humor ratings than true headlines ($M = 1.65$, 95% CI [1.58, 1.71]), $t(199) = 16.12$, $p < .001$, $d = 1.14$. Thus, false headlines were perceived as more humorous than true headlines.

6 | GENERAL DISCUSSION

In all four experiments, we found that participants rated true headlines as more accurate than false headlines. In Experiment 1, we found

that headlines paired with pictures were perceived as more accurate than headlines without. These results are consistent with findings that pictures increase the believability of general knowledge claims (Newman et al., 2012), and that pictures can create false memories in participants (Frenda et al., 2013; Lindsay et al., 2004; Wade et al., 2002). The lack of an interaction in Experiment 1 revealed that pictures influenced the perceived accuracy of true and false headlines similarly. Thus, pictures did not make people better able to discern true from false headlines, they made all headlines appear more accurate.

In addition, our results from Experiment 1 are consistent with a similar study by Strange et al. (2011). In their study, they found that pictures increased memory rates for true and false events that had never occurred. Our study adds important contributions building on Strange et al.'s (2011) work for 3 reasons. First, in Strange et al.'s (2011) investigation they state that they sourced their true headlines from "national and international news websites." However, it is not clear where their false headlines came from, if they were created by the researchers themselves, or if they were false headlines that had actually circulated virally online. Because Snopes.com fact checks stories that have circulated online and our false stories were sourced from Snopes.com, the level of ecological validity of our study importantly builds off of Strange et al. (2011). Second, Strange et al. (2011) used only 2 false headlines, whereas our study included a total of 14. Third, our study included a different outcome variable compared to Strange et al. (2011). Specifically, Strange and colleagues were interested in *memory* for news overall and *false memories* for false news. However, the present study addressed *perceived accuracy* for true and false news, and because prior literature suggests a disassociation between memory and belief (Otgaar et al., 2014), we believe this study adds important nuance and specificity to the literature.

In Experiments 2A and 2B, we found that manipulating perceptual clarity did not have the predicted effect on perceived accuracy. Perceptually clear headlines did not receive greater accuracy ratings than perceptually degraded headlines. In addition, contrary to the prediction derived from Alter et al. (2007), we found that perceptually degraded headlines did not increase discernment between true and false headlines, which should have occurred if they increased analytic thinking. A similar result was obtained by Thompson et al. (2013). They found that perceptual fluency did increase the time the participants spent thinking about the problems, but it did not have any effect on the accuracy of judgments.

In Experiment 3, we examined the effect of prior exposure on accuracy judgments. Using a different initial encoding rating task (humor) and a different set of news headlines than Pennycook et al. (2018), we replicated Pennycook et al.'s (2018) finding that prior exposure increases perceive accuracy. These results are also consistent with previous literature on the illusory truth effect (Fazio et al., 2015; Hasher et al., 1977). The lack of an interaction in Experiment 3 revealed that prior exposure influenced the perceived accuracy of true and false headlines similarly. Thus, prior exposure, like pictures, did not make people better able to discern true from false headlines, it made all headlines appear more accurate. As Pennycook

et al. (2018) pointed out, providing corrections to fake news could backfire if these corrections make the news more familiar, a phenomenon termed the *familiarity backfire effect* (Schwarz, Sanna, Skurnik, & Yoon, 2007). Future research should examine this possibility.

Another contribution of Experiment 3 is the detection of a variability in the magnitude of the repeated exposure effect. Specifically, the effect in Experiment 3 when initial ratings were for *humor* ($\eta_p^2 = .20$) was larger than the effect for Pennycook et al.'s (2018) *willingness to share* ratings ($\eta_p^2 = .09$). This suggests that the repeated exposure effect may be able to be augmented at initial encoding. Future research should explore possible alternative encoding tasks that may reduce the magnitude of the illusory truth effect.

From these results, we conclude that pictures and prior exposure seem to have a marked influence in increasing perceptions of truth. Fenn et al. (2013) claimed that pictures can increase the ease of access and familiarity to related information in memory and correspondingly make that information feel truer. We postulate that pictures (in Experiment 1) and prior exposure (in Experiment 3) may have had similar mechanisms for their effects, however because we did not manipulate familiarity directly it is possible that other factors were involved in influencing our results. The existence of a dynamic system of various contributing causes is consistent with theoretical literature on repetition induced truth effects (Unkelbach, Koch, Silva, & Garcia-Marques, 2019; Unkelbach & Stahl, 2009).

Our attempts to replicate Bronstein et al.'s (2019) and Pennycook and Rand's (2019) finding on the role of cognitive reflection in headline discernment received mixed support. Although we did not find a significant correlation between CRT and news discernibility in Experiment 1, we did in Experiment 3. Interestingly, however, we did replicate Pennycook and Rand's (2019) finding that the discernment correlation with CRT was influenced mostly by the negative correlation between CRT and false headlines. One potential reason for our null finding in Experiment 1 was that our sample consisted of college students, whereas Experiment 3, Bronstein et al. (2019) and Pennycook and Rand (2019) had Mechanical Turk workers as participants. In other studies in our lab, we have found that Mechanical Turk workers perform much better on cognitive reflection tests than our college students. These floor effects for our student sample (Experiment 1) may have reduced variability of CRT performance and masked the relationship CRT performance and discernment. In addition, we used different CRT items in Experiments 1 and 3 so no meaningful comparisons of performance can be made. We also found that CRT had poor internal consistency in Experiment 1 (Cronbach's alpha = .53), but it was acceptable in Experiment 3 (Cronbach's alpha = .72). The poor internal consistency could also explain why it did not significantly predict discernment in Experiment 1. Further research is needed to examine boundary conditions for the relationship between cognitive reflection and headline discernment.

The present study had limitations. First, we only used headlines and not actual news articles. Even true headlines are sometimes misleading of the actual content of the article (Ecker, Lewandowsky, Chang, & Pillai, 2014). Future studies can examine perceptions of accuracy for news articles rather than headlines. In addition, our

samples were also limited to college students (Experiments 1 and 2A) and Mechanical Turk (Experiments 2B and 3). Future attempts to obtain more representative samples could increase the generalizability of these findings.

In Experiment 1, an aspect that we did not control for was the probative value that each headline picture provided. Some of the pictures were more helpful than others in determining whether the headlines were real. We examined which pictures were most influential on accuracy ratings. This exploratory analysis is available in the Data S1. Based on this analysis, our participants seemed to be taking into account the probative value of the picture in their evaluations of perceived accuracy, and we have examples of this occurring for both true and false headlines. Future studies should examine this more specifically to determine to what extent probative photos matter in news discernment tasks.

Another difficulty we faced in our study, and of fake news studies in general, is the process of selecting news headlines. For true headlines, because researchers want to obtain variability in their outcome believability measures, they do not usually include blatantly obvious headlines (e.g., who won the Superbowl or who won the 2016 presidential election). However, this can lead to ecological validity problems if too many surprising and true headlines are selected for their samples. We searched for our headlines keeping these issues in mind, attempting to select headlines so that the surprising, true headlines did not dominate our sample. However, it is important to recognize that we did not control for the surprising nature of the headlines. We also did not control for the humorous aspect of headlines. Although, false news does seem to have a tendency towards humor (Tandoc Jr, Lim, & Ling, 2018), given the nature of our Experiment 3, it was critical that we were aware of any issues with selection biases, especially as it relates to the humor of our fake news stimuli. Nonetheless, it is possible that we may have been biased in our selection of headlines (in order to obtain plausible false and implausible true headlines).

Regarding issues of generalizability, we want to conclude with three points. Despite the fact that we attempted to select headlines in a way that could generalize to a broader category of news, we acknowledge the difficulty of the task and that our set may lack strong generalizability. However, addressing Experiment 1, including 14 false headlines from Snopes compared to Strange et al.'s (2011) two from unknown sources still adds a meaningful level of external validity. In all of our experiments, we attempted to create our stimuli by mimicking the typical presentation of news information on social media and online – an image accompanying a few lines of text. Therefore, we believe they have strong ecological validity.

According to a common notion suggesting that humans are completely rational beings, we should – therefore – be able to make objective decisions based on reading news headlines, free from influences such as visual stimuli, perceptual clarity, and repeated exposure. However, our results suggest that visual information (i.e., images) and repeated exposure have a pervasive effect on accuracy judgments of headlines, similar to the effects that pictures (e.g., Newman et al., 2012) and prior exposure (e.g., Fazio et al., 2015; Hasher et al., 1977; Pennycook et al., 2018) have had in other tasks.

Our finding that people rated true headlines as more accurate than false headlines is also consistent with those from other fake news studies (Bronstein et al., 2019; Pennycook & Rand, 2019), which suggests that people are able to discern true from false headlines. However, it is important to note that given a different sampling approach, a different result could be obtained. Therefore, we conclude that regarding our set of materials, participants were able to discern true from false. Interestingly, although we consistently found that people were able to discern true from false headlines, the differences were not particularly large: in Experiment 1, mean differences between true and false headlines were 0.12; in Experiment 2A, they were 0.58; in Experiment 2B, they were 0.94; and in Experiment 3, they were 0.59 (on a 4-point scale; effect sizes of η_p^2 = .10, .61, .77, and .64, respectively). These small differences thus illustrate the need for future research that examines the efficacy of training techniques to improve news discernment.

This study explored some of the reasons why people have been shown to be susceptible to fake news. Like in Vosoughi et al.'s (2018) Twitter study and Silverman et al.'s (2016) Facebook study, there is evidence that fake news poses a serious threat and warrants our concern and that social media is contributing to the problem. Because the impetus of these social media entities is to encourage the sharing of content, of which is frequently accompanied by visual stimuli (e.g., pictures), understanding the influence of repetition and pictures on truth evaluations for news is paramount. This study contributes to our understanding of how the impact of pictures and repeated exposure may be influencing this current trend of fake news proliferation and suggests that factors that seem to increase familiarity may increase perceived accuracy. Being aware that we can be unintentionally influenced by these stimuli is the first step towards trying to ameliorate the problem.

CONFLICT OF INTEREST

The authors declare no conflicts of interest related to this research.

ETHICAL STATEMENT

All experiments described in this manuscript were approved by an Institutional Review Board prior to data collection. All participants provided informed consent prior to their inclusions in a study, and all agreed to having their data available on the OSF.

DATA AVAILABILITY STATEMENT

The hypotheses, data collection plans, and planned analyses for all experiments described in this manuscript were preregistered on the Open Science Framework (OSF). Additionally, the stimuli and data from each experiment are available on the OSF. Experiment 1 information is available at https://osf.io/fq9vr/, Experiment 2a and 2b at https://osf.io/jqxa2/, and Experiment 3 at https://osf.io/8wbqu/. We

note all deviations from our preregistered plans in the the manuscript and supplementary material.

ORCID

Dustin P. Calvillo https://orcid.org/0000-0002-6747-9605

REFERENCES

Abed, E., Fenn, E., & Pezdek, K. (2017). Photographs elevate truth judgments about less well-known people (but not yourself). *Journal of Applied Research in Memory and Cognition*, 6, 203–209. https://doi.org/10.1016/j.jarmac.2017.01.007

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569–576. https://doi.org/10.1037/0096-3445.136.4.569

Baron, J., Scott, S., Fincher, K. S., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, 4, 265–284. https://doi.org/10.1016/j.jarmac.2014.09.003

Bronstein, M. V., Pennycook, G., Bear, A., Rand, D. G., & Cannon, T. D. (2019). Belief in fake news is associated with delusionality, dogmatism, religious fundamentalism, and reduced analytic thinking. *Journal of Applied Research in Memory and Cognition*, 8(1), 108–117. https://doi.org/10.1016/j.jarmac.2018.09.005

Cardwell, B. A., Lindsay, D. S., Förster, K., & Garry, M. (2017). Uninformative photos can increase people's perceived knowledge of complicated processes. *Journal of Applied Research in Memory and Cognition*, 6, 244–252. https://doi.org/10.1016/j.jarmac.2017.05.002

Ecker, U. K., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied*, 20(4), 323–335. https://doi.org/10.1037/xap0000028

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

Fenn, E., Newman, E. J., Pezdek, K., & Garry, M. (2013). The effect of nonprobative photographs on truthiness persists over time. *Acta Psychologica*, 144, 207–211. https://doi.org/10.1016/j.actpsy.2013.06.004

Frenda, S. J., Knowles, E. D., Saletan, W., & Loftus, E. F. (2013). False memories of fabricated political events. *Journal of Experimental Social Psychology*, 49, 280–286. https://doi.org/10.1016/j.jesp.2012.10.013

Gervais, W. M., & Norenzayan, A. (2012). Analytic thinking promotes religious disbelief. *Science*, 336, 493–496. https://doi.org/10.1126/science.1215647

Graesser, A. C., & Hemphill, D. (1991). Question answering in the context of scientific mechanisms. *Journal of Memory and Language*, 30(2), 186–209.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112.

Henkel, L. A. (2011). Photograph-induced memory errors: When photographs make people claim they have done things they have not. *Applied Cognitive Psychology*, 25, 78–86. https://doi.org/10.1002/acp.1644

Hsu, S. (2017, March 24). Comet Pizza gunman pleads guilty to federal and local charges. *The Washington Post*, Retrieved from https://www.washingtonpost.com/local/public-safety/comet-pizza-gunman-to-appear-at-plea-deal-hearing-friday-morning/2017/03/23/e12c91ba-0986-11e7-b77c-0047d15a24e0_story.html

Kang, C. (2016, November 21). Fake news onslaught targets pizzeria as nest of child-trafficking. *New York Times*, Retrieved from https://www.nytimes.com/2016/11/21/technology/fact-check-this-pizzeria-is-not-a-child-trafficking-site.html

Kang, C., & Goldman, A. (2016, December 5). In Washington pizzeria attack, fake news brought real guns. *New York Times*, Retrieved from https://www.nytimes.com/2016/12/05/business/media/comet-ping-pong-pizza-shooting-fake-news-consequences.html

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., ... Batra, R. (2018). Many labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1, 443–490. https://doi.org/10.1177/2515245918810225

LaCapria, K. (2016, November 21). Is comet ping pong pizzeria home to a child abuse ring led by Hillary Clinton? *Snopes.com*, Retrieved from https://www.snopes.com/fact-check/pizzagate-conspiracy/

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... Schudson, M. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lindsay, D. S., Hagen, L., Read, J. D., Wade, K. A., & Garry, M. (2004). True photographs and false memories. *Psychological Science*, 15, 149–154. https://doi.org/10.1111/j.0956-7976.2004.01503002.x

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, 16–30. https://doi.org/10.1037/xge0000049

Mitchell, A., Gottfried, J., Stocking, G., Walker, M., & Fedeli, S. (2019). Many Americans say made-up news is a critical problem that needs to be fixed. Pew Research Center, Retrieved from https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/

Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, 19, 969–974. https://doi.org/10.3758/s13423-012-0292-0

Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D. M., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 1337–1348. https://doi.org/10.1037/xlm0000099

Oldrati, V., Patricelli, J., Colombo, B., & Antonietti, A. (2016). The role of dorsolateral prefrontal cortex in inhibition mechanism: A study on cognitive reflection test and similar tasks through neuromodulation. *Neuropsychologia*, 91, 499–508. https://doi.org/10.1016/j.neuropsychologia.2016.09.010

Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science*, 15, 100–105. https://doi.org/10.1111/j.0963-7214.2004.01502005.x

Otgaar, H., Scoboria, A., & Mazzoni, G. (2014). On the existence and implications of nonbelieved memories. *Current Directions in Psychological Science*, 23(5), 349–354. https://doi.org/10.1177/0963721414542102

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147, 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 20(2), 185–200. https://doi.org/10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, 29(5), 453–469. https://doi.org/10.1002/bdm.1883

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, 8(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*, 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology*, *39*, 127–161. https://doi.org/10.1016/S0065-2601(06)39003-X

Silverman, C., Strapagiel, L., Shaban, H., & Hall, E. (2016). Hyperpartisan Facebook pages are publishing false and misleading information at an alarming rate. *Buzzfeed News*, Retrieved from https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis

Song, H., & Schwarz, N. (2008a). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, *19*, 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Song, H., & Schwarz, N. (2008b). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, *26*, 791–799. https://doi.org/10.1521/soco.2008.26.6.791

Strange, D., Garry, M., Bernstein, D. M., & Lindsay, D. S. (2011). Photographs cause false memories for the news. *Acta Psychologica*, *136*, 90–94. https://doi.org/10.1016/j.actpsy.2010.10.006

Tandoc, E. C., Jr., Lim, Z. W., & Ling, R. (2018). Defining "fake news" a typology of scholarly definitions. *Digital Journalism*, *6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Thompson, V., Turner, J., Pennycook, G., Ball, L., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*, 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*, 99–113.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, *20*(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Topolinski, S., & Strack, F. (2010). False fame prevented: Avoiding fluency effects without judgmental correction. *Journal of Personality and Social Psychology*, *98*, 721–733. https://doi.org/10.1037/a0019260

Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, *28*(3), 247–253. https://doi.org/10.1177/0963721419827854

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

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wade, K. A., Garry, M., Read, J. D., & Lindsay, D. S. (2002). A picture is worth a thousand lies: Using false photographs to create false childhood memories. *Psychonomic Bulletin & Review*, *9*, 597–603. https://doi.org/10.3758/BF03196318

SUPPORTING INFORMATION

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