RUNNING HEAD: PRIOR EXPOSURE AND FAKE NEWS

Implausibility and illusory truth: Prior exposure increases perceived accuracy of fake news but has no effect on entirely implausible statements

Gordon Pennycook[1]*, Tyrone D. Cannon[1], & David G. Rand[1,2,3]

[1]Department of Psychology, [2]Department of Economics, [3]School of Management, Yale University, 1 Prospect Street, New Haven, CT 06511, USA

*Corresponding author: gordon.pennycook@yale.edu

This manuscript has not been peer-reviewed.

First version upload date: April 30, 2017

Second version upload date: August 26, 2017

This version upload date: December 11, 2017

Key Words: fake news; news media; social media; fluency; illusory truth effect

Abstract

The 2016 US Presidential Election brought considerable attention to the phenomenon of "fake news": entirely fabricated and often partisan content that is presented as factual. Here we demonstrate one mechanism that contributes to the believability of fake news: prior exposure. Using actual fake news headlines presented as they were seen on Facebook, we show that even a single exposure increases subsequent perceptions of accuracy, both within the same session and after a week – that is, an illusory truth effect exists for fake news. Moreover, increased perceptions of accuracy for repeated fake news headlines occurs even when the stories are labeled as contested by fact checkers, or are inconsistent with the reader's political ideology. These results suggest that social media platforms help to incubate blatantly false news stories, and that tagging such stories as disputed is not an effective solution to this problem. Interestingly, however, we also find that prior exposure does *not* impact entirely implausible statements, and provide evidence for an inverted U-shaped relationship between plausibility and the magnitude of the illusory truth effect. These observations support of a model of the illusory truth effect in which fluency and prior knowledge interact, rather than proceeding serially.

The ability to form accurate beliefs, particularly about issues of great importance, is key to our success as individuals as well as the functioning of our societal institutions (and, in particular, democracy). Across a wide range of domains, it is critically important to correctly assess what is true and what is false: Accordingly, differentiating real from unreal is at the heart of our societal constructs of rationality and sanity (Corlett, 2009; Sanford, Veckenstedt, & Moritz, 2014). Yet the ability to form and update beliefs about the world sometimes goes awry – and not just in the context of inconsequential, small-stakes decisions.

The potential for systematic inaccuracy in important beliefs has been particularly highlighted by the wide-spread consumption of disinformation during the 2016 US Presidential Election. This is most notably exemplified by so-called "fake news" – that is, news stories that were fabricated (but presented as if from legitimate sources) and promoted on social media in order to deceive the public for ideological and/or financial gain. An analysis of the top performing news articles on Facebook in the months leading up to the election revealed that the top fake news articles actually outperformed the top real news articles in terms of shares, likes, and comments (Silverman, Strapagiel, Shaban, & Hall, 2016). Although it is unclear to what extent fake news influenced the outcome of the Presidential Election (Allcott & Gentzkow, 2017), there is no question that many people were deceived by entirely fabricated (and often quite fanciful) fake news stories – including, for example, high-ranking government officials, such as Pakistan's defense minister (Goldman, 2016). How is it that so many people came to believe stories that were patently and demonstrably untrue? What mechanisms underlie these false beliefs that might be called mass delusions?

Here, we explore one potential answer: prior exposure. Given the ease with which fake news can be created and distributed on social media platforms (Shane, 2017), combined with our increasing tendency to consume news via social media (Gottfried & Shearer, 2016), it is likely that we are being exposed to fake news stories with much greater frequency than in the past. Might exposure *per se* help to explain people's tendency to believe outlandish political disinformation?

The Illusory Truth Effect

There is a long tradition of work in cognitive science demonstrating that prior exposure to a statement (e.g., "The capybara is the largest of the marsupials") increases the likelihood that participants will judge it to be accurate (Arkes, Boehm, & Xu, 1991; Bacon, 1979; Begg, Anas, & Farinacci, 1992; Dechene, Stahl, Hansen, & Wanke, 2010; Fazio, Brashier, Payne, & Marsh, 2015; Hasher, Goldstein, & Toppino, 1977; Polage, 2012; Schwartz, 1982). The dominant account of this "illusory truth effect" is that repetition increases the ease with which statements are processed (i.e., processing fluency), which in turn is used heuristically to infer accuracy (Alter & Oppenheimer, 2009; Begg et al., 1992; Reber, Winkielman, & Schwarz, 1998; Unkelbach, 2007; Wang, Brashier, Wing, Marsh, & Cabeza, 2016; Whittlesea, 1993, but see Unkelbach & Rom, 2017). Past studies have shown this phenomenon using a range of innocuous and plausible statements, such as obscure trivia questions (Bacon, 1979) or assertions about consumer products (Hawkins & Hoch, 1992; Johar & Roggeveen, 2007). Repetition can even increase the perceived accuracy of plausible but false statements among participants who are subsequently able to identify the correct answer (Fazio et al., 2015).

Here we ask whether illusory truth effects will extend to fake news. Given that the fake news stories circulating on social media are quite different from the stimuli that have been employed in previous illusory truth experiments, finding such an effect for implausible and highly partisan fake news extends the scope (and real-world relevance) of the effect and, as we will argue, informs theoretical models of the effect. Indeed, there are numerous reasons to think that simple prior exposure will not extend to fake news.

Implausibility as a potential boundary condition of the illusory truth effect

Fake news stories are constructed with the goal of drawing attention, and are therefore often quite fantastical and implausible. For example, Pennycook and Rand (2017a) gave participants a set of politically partisan fake news headlines collected from online websites (e.g., "Trump to Ban All TV Shows that Promote Gay Activity Starting with Empire as President"), and found that they were only judged as accurate 17.8% of the time. To contrast this figure with the existing illusory truth literature, Fazio et al. (2015) found that false trivia items were judged to be true around 40% of the time, even when restricting the analysis to participants who were subsequently able to recognize the statement as false. Thus, these previous statements (such as "chemosynthesis is the name of the process by which plants make their food"), despite being untrue, are much more plausible than typical fake news headlines. This may have consequences for whether repetition increases perceived accuracy of fake news: When it is completely obvious that a statement is false, it may be perceived as inaccurate regardless of how fluently it is processed. Although such an influence of plausibility is not explicitly part of the Fluency-Conditional Model of illusory truth proposed by Fazio and colleagues (under which knowledge only influences judgment when people do not rely on fluency), the possibility of such an effect is acknowledged in their discussion when they state that they "expect that participants would draw on their knowledge, regardless of fluency, if statements contained implausible errors."

Motivated reasoning as a potential boundary condition of the illusory truth effect

Another striking feature of fake news that may counteract the effect of repetition – and which is absent from prior studies of the illusory truth effect – is the fact that fake news stories are not only political in nature, but are often extremely partisan. Although prior work has shown the illusory truth effect *on average* for (relatively innocuous) social-political opinion statements (Arkes, Hackett, & Boehm, 1989), the role of individual differences in ideological discordance has not been examined. Importantly, people have a strong motivation to reject the veracity of stories that conflict with their political ideology (Flynn, Nyhan, & Reifler, 2016; Kahan, 2013; Kahan et al., 2012), and the hyper-partisan nature of fake news makes such conflicts virtually assured for roughly half the population. Furthermore, the fact that fake news stories are typically of immediate real-world relevance – and therefore, presumably, more impactful on a person's beliefs and actions than the relatively trivial pieces of information considered in previous work on the illusory truth effect – should make people more inclined to think carefully about the accuracy of such stories, rather than relying on simple heuristics when making accuracy judgments. Thus, there is reason to expect that people may be resistant to illusory truth effects for partisan fake news stories that they have politically motivated reasons to reject.

The current work

Although there are many reasons that, in theory, people should not believe fake news (even if they have seen it before), it is clear that many people do in fact find such stories credible. Fake news is a topic of great general interest for this very reason. If repetition increases perceptions of accuracy even for highly implausible and partisan content, then increased exposure may (at least partly) explain why fake news stories have recently proliferated. Here we assess this possibility with a set of highly powered and preregistered experiments. In a first study, we explore the impact of extreme implausibility on the illusory truth effect in the context of politically neutral statements. We find that implausibility does indeed present a boundary condition for illusory truth, such that repetition does not increase perceived accuracy of statements which essentially no one believes at baseline. In two more studies, however, we find that – despite being implausible, partisan, and provocative – fake news headlines that are repeated are in fact perceived as more accurate. In a final study, we inform theoretical models of the illusory truth effect by quantitatively assessing the link between (im)plausibility and illusory truth using the pooled results of our three experiments. Doing so, we find evidence for an inverted-U shaped relationship between baseline plausibility and the magnitude of the illusory truth effect: although extremely implausible statements appear to be immune to repetition, small initial increases in plausibility are associated with increasing illusory truth effects; furthermore, we find that the effect size begins to decrease around the mid-point on our plausibility scale. Taken together, these results shed light on how people come to have patently false beliefs, help to inform efforts to reduce such beliefs, and extend our theoretical understanding of the basis of illusory truth effects.

Study 1 – Extreme Implausibility Boundary Condition

Although existing models of the illusory truth effect do not explicitly take plausibility into account, we hypothesized that prior exposure should *not* increase perceptions of accuracy for statements that are *prima facie* implausible. That is, when strong internal reasons exist to reject the veracity of a statement, it should not matter how fluently the statement is processed.

To assess implausibility as a boundary condition for the illusory truth effect, we created statements that participants would certainly *know* to be false (i.e., extremely implausible statements such as “The Earth is a perfect square”) and manipulated prior exposure using a standard illusory truth paradigm (via Fazio et al., 2015). We also included unknown (but plausible) true and false trivia statements from a set of general knowledge norms (Tauber, Dunlosky, & Rawson, 2013). To balance out the set, participants were also given easily known truths (see Table 1 for example items from each set). Participants first rated the "interestingness" of half of the items and, following an unrelated intervening questionnaire, they were asked to assess the accuracy of all items. Thus, half of the items in the assessment stage were previously presented (i.e., familiarized) and half were novel. If implausibility is a boundary condition for the illusory truth effect, there should be no significant effect of repetition on extremely implausible (known) falsehoods. We expect to replicate the standard illusory truth effect for unknown (but plausible) trivia statements. For extremely plausible known true statements, there may be a ceiling effect on accuracy judgments that precludes an effect of repetition (c.f. results for fluency on known truths, Unkelbach, 2007).

Table 1. Example items from Study 1.

Known	True	There are more than fifty stars in the universe.
	False (Implausible)	The earth is a perfect square.
Unknown	True	Billy the Kid's last name was Bonney.
	False	Angel Falls is located in Brazil.

Method

All data are available online (https://osf.io/txf46/). We preregistered our hypotheses, primary analyses, and sample size (https://osf.io/txf46/). Although one-tailed tests are justified in the case of pre-registered directional hypotheses, here we follow conventional practices and use two-tailed tests throughout (the use of one-tailed versus two-tailed tests does not qualitatively alter our results). All participants were recruited from Amazon's Mechanical Turk (Horton, Rand, & Zeckhauser, 2011), which has been shown to be a reliable resource for research on political ideology (Coppock, 2016; Krupnikov & Levine, 2014; Mullinix, Leeper, Druckman, & Freese, 2015). These studies were approved by the Yale Human Subject Committee.

Participants. Our target sample was 500. In total, 566 participants completed some portion of the study. We had complete data for 515 participants (51 participants dropped out). Participants were removed if they indicated responding randomly (N = 50), or searching online for any of the claims (N = 24; 1 of which did not respond), or going through the familiarization stage without doing the task (N = 32). These exclusions were preregistered. The final sample (N = 409; Mean age = 35.8) included 171 males and 235 females (3 did not indicate sex).

Materials. We created 4 known falsehoods (i.e., extremely implausible statements) and 4 known truths statements (see Supplementary Information, SI, for full list). We also used 10 true and 10 false trivia questions framed as statements (via Tauber, Dunlosky, & Rawson, 2013). Trivia items were sampled from largely unknown facts (see Table 1).

Procedure. We used a parallel procedure to Fazio et al. (2015). Participants were first asked to rate the "interestingness" of the items on a 6 point scale from 1) very uninteresting to 6) very interesting. Half of the items were presented in this familiarization stage (counterbalanced). Participants then completed a few demographic questions and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). This filler stage consisted of 25 questions and took approximately two minutes. Demographic questions consisted of age ("What is your age?"), sex ("What is your sex?"), education ("What is the highest level of school you have completed or the highest degree you have received" with 8 typical education level options), English fluency ("Are you fluent in English"), and zip code ("Please enter the ZIP code for your primary residence. Reminder: This survey is anonymous"). Finally, participants were asked to assess the accuracy of the statements on a 6 point scale from 1) definitely false to 6) definitely true. At the end of the survey, participants were asked about random responding (Did you respond randomly at any point during the study?) and use of search engines (Did you search the internet (via Google or otherwise) for any of the news headlines?). Both were accompanied by a "yes/no" response option and the following clarification: "Note: Please be honest! You will get your HIT regardless of your response."

Results

Following our preregistration, the key comparison was between familiarized and novel implausible items. As predicted, repetition did not increase perceptions of accuracy for implausible (known false) statements, *p* = .462 (see Table 2), while there was a significant effect of repetition for both true and false trivia (unknown) statements, *p*'s < .001. There was no significant effect of repetition on very plausible (known true) statements (*p* = .078). These results were supported by a significant interaction between knowledge (known, unknown) and exposure (familiarized, novel), $F(1, 408) = 82.17$, $MSE = .35$, $p < .001$, $\eta^2 = .17$. Specifically, there was no significant overall effect of repetition for known items, $F(1, 408) = .91$, $MSE = .30$, $p = .341$, $\eta^2 = .002$, but a highly significant overall effect for unknown items, $F(1, 408) = 107.99$, $MSE = .47$, $p < .001$, $\eta^2 = .21$.

Table 2. Means, standard deviations, and significance tests (comparing familiarized and novel items) for known or unknown true and false statements in Study 1.

	Type	Familiarized	Novel	*t* (df)	*p*
Known	True	5.59 (0.8)	5.66 (0.6)	1.77 (408)	.078
	False (Implausible)	1.13 (0.6)	1.11 (0.5)	0.74 (408)	.462
Unknown	True	4.12 (0.7)	3.79 (0.8)	6.65 (408)	< .001
	False	3.77 (0.7)	3.39 (0.7)	9.44 (408)	< .001

Discussion

While we replicated prior results indicating a positive effect of familiarity on ambiguously plausible statements, regardless of their correctness, we observed *no* significant effect of familiarity on accuracy judgements for statements which are patently false.

Study 2 – Fake News

Study 1 establishes that, at least, *extreme* implausibility is a boundary condition for the illusory truth effect. Nonetheless, given that fake news stories are highly (but not extremely) implausible (Pennycook & Rand, 2017a), it is unclear whether their level of plausibility will be sufficient to allow prior exposure to inflate the perceived accuracy of fake news. It is also unclear what impact the highly partisan nature of fake news stimuli, and the motivated reasoning to which this partisanship may lead (i.e., reasoning biased toward conclusions that are concordant with previous opinion; Kahan, 2013; Kunda, 1990; Mercier & Sperber, 2011; Redlawsk, 2002), will have on any potential illusory truth effect. Motivated reasoning may create a motivation to see politically discordant stories as disproportionally inaccurate, such that the illusory truth effect may be diluted (or reversed) when headlines are discordant. We assess these questions in Study 2.

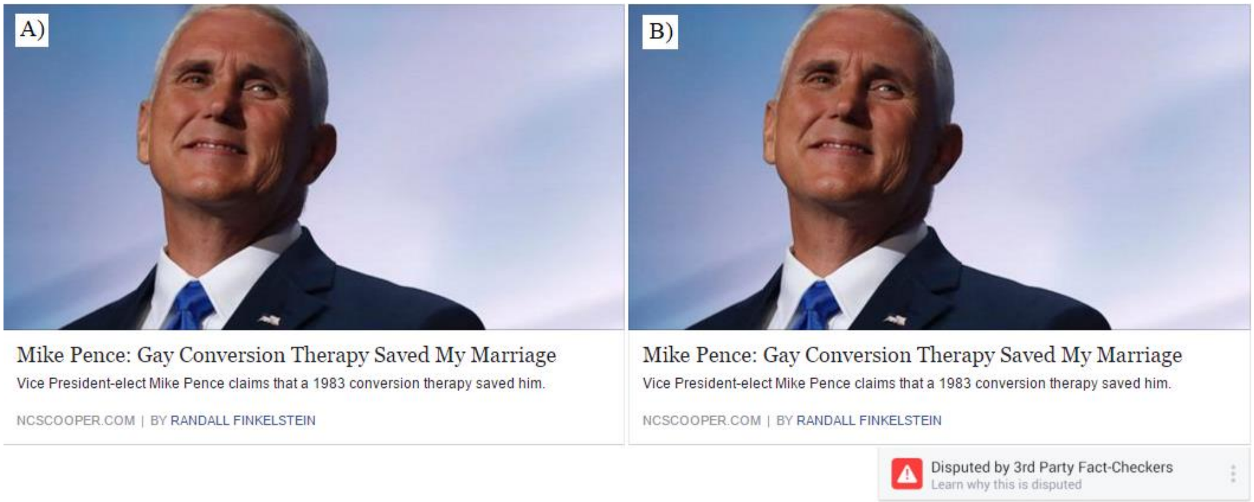
In addition to assessing the baseline impact of repetition on fake news, we also investigated the impact of explicit warnings about a lack of veracity on the illusory truth effect, given that warnings have been shown to be effective tools for diminishing (although not abolishing) the memorial effects of misinformation (Ecker, Lewandowsky, & Tang, 2010). Furthermore, such warnings are a key part of efforts to combat fake news – for example, Facebook presently flags stories shown to be false with a caution symbol and the text "Disputed by 3rd Party Fact-Checkers" (Mosseri, 2016). To this end, half of the participants were randomly assigned to a Warning condition in which this caution symbol and "Disputed" warning were applied to the fake news headlines. Prior work has shown that participants rate repeated trivia statements as more accurate than novel statements, even when they were told that the source was inaccurate (Begg et al., 1992). Specifically, Begg and colleagues attributed statements in the familiarization stage to people with either male or female names, and then told participants that either all males or all females were lying. Participants were then presented with repeated and novel statements – all without sources – and they rated previously presented statements as more accurate even if they had been attributed to the lying gender in the familiarization stage. This provides evidence that the illusory truth effect survives manipulations that decrease belief in statements at first exposure. Nonetheless, Begg and colleagues employed a design in which information about veracity was provided indirectly: for any given statement presented during their familiarization phase, participants had to complete the additional step at encoding of mapping the source's gender into the information provided about which gender was unreliable in order to inform their initial judgment about accuracy. The "Disputed" warnings we test here, conversely, are more direct and do not involve this extra mapping step. Thus, by assessing their impact on the illusory truth effect, we test the robustness of Begg and colleagues' findings to this more explicit warning, while also generating practically useful insight into the efficacy of this specific fake news intervention.

Method

Participants. We had an original target sample of 500 participants in our preregistration. We then completed a full replication of the experiment with another 500 participants. Given the similarity across the two samples, the datasets were combined for the main analysis (the results are qualitatively similar when examining the two experiments separately, see SI). The first wave was completed on January 16th and the second wave was completed on February 3rd (both in 2017). In total, 1069 participants from Mechanical Turk completed some portion of the survey. However, 64 did not finish the study and were removed (33 from the no warning condition and 31 from the warning condition). A further 32 participants indicated responding randomly at some point during the study and were removed. We also removed participants who reported searching for the headlines (N = 18) or skipping through the familiarization stage (N = 6). These exclusions were preregistered for Studies 1 and 3, but accidentally omitted from the preregistration for Study 2. The results are qualitatively identical with the full sample, but we report analyses with participants removed to retain consistency across our studies. The final sample (N = 949; Mean age = 37.1) included 449 males and 489 females (11 did not respond).

Materials and Procedure. Participants engaged in a 3-stage experiment. In the familiarization stage, participants were shown six news headlines that were factually accurate (*real news*) and six others that were entirely untrue (*fake news*). The headlines were presented in an identical format to that of Facebook posts (i.e., a headline with an associated photograph above it and a byline below it; see Figure 1a). Participants were randomized into two conditions: 1) The warning condition where all of the fake news headlines (but none of the real news headlines) in the familiarization stage were accompanied by a “Disputed by 3rd Party Fact-Checkers” tag” (see Figure 1b), or 2) The control condition where fake and real news headlines were displayed without warnings. In the familiarization stage, participants engaged with the news headlines in an ecologically valid way: they indicated whether they would share each headline on social media. Specifically, participants were asked “Would you consider sharing this story online (for example, through Facebook or Twitter)?” and were given three response options (“No, Maybe, Yes”). For purposes of data analysis, “no” was coded as 0 and “maybe” and “yes” were coded as 1[1].

[1] This was not preregistered for Study 2; however, it was for Study 3. Hence, we use this analysis strategy to retain consistency across the two fake news studies. The results are qualitatively similar if the social media question is scored continuously.



Mike Pence: Gay Conversion Therapy Saved My Marriage
Vice President-elect Mike Pence claims that a 1983 conversion therapy saved him.
NCSCOOPER.COM | BY RANDALL FINKELSTEIN

Mike Pence: Gay Conversion Therapy Saved My Marriage
Vice President-elect Mike Pence claims that a 1983 conversion therapy saved him.
NCSCOOPER.COM | BY RANDALL FINKELSTEIN

Disputed by 3rd Party Fact-Checkers
Learn why this is disputed

Figure 1. *Sample fake news headline without (A) and with (B) "Disputed" warning, as presented in Experiments 2 and 3.*

The participants then advanced to the distractor stage, in which they completed a set of filler demographic questions. These included: age, sex, education, proficiency in English, political party (Democratic, Republican, Independent, other), social and economic conservatism (separate items)[2], and two questions about the 2016 election. For these election-related questions, participants were first asked to indicate who they voted for (given the following options: Hillary Clinton, Donald Trump, Other Candidate (such as Jill Stein or Gary Johnson), I did not vote for reasons outside my control, I did not vote but I could have, and I did not vote out of protest). Participants were then asked "If you absolutely had to choose between only Clinton and Trump, who would you prefer to be the next President of the United States". This binary was then used as our political ideology variable for the concordance/discordance analysis. Specifically, for participants who indicated a preference for Trump, Pro-Republican stories were scored as politically concordant and Pro-Democrat stories were scored as politically discordant; for

[2] Participants were asked "On social issues I am: Strongly Liberal, Somewhat Liberal, Moderate, Somewhat Conservative, Strongly Conservative". The same was true for the economic conservatism item except the prompt was "On economic issues I am:".

participants who indicated a preference for Clinton, Pro-Democrat stories were scored as politically concordant and Pro-Republican stories were scored as politically discordant. The filler stage took approximately one minute.

Finally, participants entered the assessment stage, where they were presented with 24 news headlines – the 12 headlines they saw in the familiarization stage and 12 new headlines (6 fake news, 6 real news) – and rated each for familiarity and accuracy. Which headlines were presented in the familiarization stage was counterbalanced across participants, and headline order was randomized for every participant in both Stage 1 and Stage 3. Moreover, the items were balanced politically, with half being Pro-Democrat and half Pro-Republican. The fake news headlines were selected from Snopes.com, a third-party website that fact-checks news stories. The real headlines were contemporary stories from mainstream news outlets. For each item, participants were first asked "Have you seen or heard about this story before?" and were given three response options ("No, Unsure, Yes"). For the purposes of data analysis, "yes" and "unsure" were combined (this analysis was not preregistered). As in other work on perceptions of news accuracy (Pennycook & Rand, 2017a; 2017b), participants were then asked "To the best of your knowledge, how accurate is the claim in the above headline?" and they rated accuracy on the following 4-point scale: 1) not at all accurate, 2) not very accurate, 3) somewhat accurate, 4) very accurate. We focus on judgments about news *headlines*, as opposed to full articles, because much of the public's engagement with news on social media involves only reading story headlines (Gabielkov, Ramachandran, & Chaintreau, 2016).

At the end of the survey, participants were asked about random responding, use of search engines the check accuracy of the stimuli, and whether they skipped through the familiarization stage ("At the beginning of the survey (when you were asked whether you would share the stories on social media), did you just skip through without reading the headlines"?). All were accompanied by a "yes/no" response option.

Our preregistration specified the comparison between familiarized and novel fake news, separately in the warning and no warning conditions, as the key analyses. However, for completeness, we report the full set of analyses that emerge from our mixed design ANOVA. Our political concordance analysis deviates somewhat from the analysis that was preregistered (see footnote 3), and our follow-up analysis that focuses on unfamiliar headlines was not preregistered. Our full preregistration is available at the following link: https://osf.io/txf46/.

Results

As a manipulation check for our familiarization procedure, we submitted familiarity ratings (recorded during the assessment stage) to a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA. Critically, there was a main effect of exposure such that familiarized headlines were rated as more familiar (M = 44.7%, SD = 35.6) than novel headlines (M = 16.2%, SD = 15.5), $F(1, 947) = 578.76$, $MSE = .13$, $p < .001$, $\eta^2 = .38$, and a significant simple effect was present within every combination of news type and warning condition, all t's > 14.0, all p's < .001. This indicates that our social media sharing task in the familiarization stage was sufficient to capture participants' attention. Further analysis of familiarity judgments can be found in supplementary information.

As a manipulation check for attentiveness to the "Disputed by 3rd party fact-checkers" warning, we submitted the willingness to share news articles on social media measure (from the familiarization stage) to a 2 (Type: fake, real) x 2 (Warning: warning, no warning) mixed design ANOVA. This analysis revealed a significant main effect of type, such that our participants were somewhat more willing to share real stories (M = 15.1%, SD = 20.8) than fake stories (M =

13.5%, *SD* = 20.8), *F*(1, 717) = 4.36, *MSE* = .02, *p* = .037, η^2 = .006. More importantly, there was a significant main effect of warning, *F*(1, 717) = 6.09, *MSE* = .06, *p* = .014, η^2 = .008, which was qualified by an interaction between type and warning, *F*(1, 717) = 10.43, *MSE* = .02, *p* = .001, η^2 = .014, such that relative to the No Warning condition, participants in the Warning condition reported being less willing to share fake news headlines (Warning: *M* = 10.2%, *SD* = 18.2; No Warning: *M* = 16.6%, *SD* = 22.6), *t*(947) = 4.78, *p* < .001, *d* = .31, whereas there was no significant warning condition effect for sharing of real news (Warning: *M* = 14.7%, *SD* = 20.9; No Warning: *M* = 15.4%, *SD* = 20.7), *t* < 1. Thus, participants clearly paid attention to the warning.

We now turn to perceived accuracy, our main focus. Perceived accuracy was entered into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA. Demonstrating the existence of an illusory truth effect, there was a significant main effect of exposure, *F*(1, 947) = 93.65, *MSE* = .12, *p* < .001, η^2 = .09, such that headlines presented in the familiarization stage (*M* = 2.24, *SD* = .42) were rated as more accurate than novel headlines (*M* = 2.13, *SD* = .39). There was also a significant main effect of headline type, such that real news headlines (*M* = 2.67, *SD* = .48) were rated as much more accurate than fake news headlines (*M* = 1.71, *SD* = .46), *F*(1, 945) = 2424.56, *MSE* = .36, *p* < .001, η^2 = .72. However, there was no significant interaction between exposure and type of news headline, *F* < 1. In particular, prior exposure increased accuracy ratings even when only considering fake news headlines (see Figure 2; Familiarized: *M* = 1.77, *SD* = .56; Novel: *M* = 1.65, *SD* = .48), *t*(948) = 7.60, *p* < .001, *d* = .25. For example, nearly twice as many participants (92.1% increase, from 38 to 73 out of 949 total) judged the fake news headlines presented to them during the familiarization stage as accurate (mean accuracy rating above 2.5), compared to the stories presented to them for the first time in the assessment stage. Although both of these participant counts are only a small fraction of the total sample, the fact that a single exposure to the fake stories doubled the number of credulous participants suggests that repetition effects may have a substantial impact in daily life, where people can see fake news headlines cycling many times through their social media newsfeeds.

What effect did the presence of warnings on fake news in the familiarization stage have on later judgments of accuracy and, potentially, the effect of repetition? The ANOVA described above revealed a significant main effect of the warning manipulation, *F*(1, 947) = 5.39, *MSE* = .53, *p* = .020, η^2 = .005, indicating that the warning decreased perceptions of news accuracy. However, this was qualified by an interaction between warning and type, *F*(1, 947) = 5.83, *MSE* = .36, *p* = .016, η^2 = .006. Whereas the presence of warnings on fake news in the assessment stage had no effect on perceptions of real news accuracy (Warning: *M* = 2.67, *SD* = .49; No Warning: *M* = 2.67, *SD* = .48), *t* < 1, participants rated fake news as less accurate in the warning condition (Warning: *M* = 1.66, *SD* = .46; No Warning: *M* = 1.76, *SD* = .46), *t*(947) = 3.40, *p* = .001, *d* = .22. Furthermore, there was a marginally significant interaction between exposure and warning, *F*(1, 947) = 3.32, *MSE* = .12, *p* = .069, η^2 = .004, such that the decrease in overall perceptions of accuracy was significant for familiarized items (Warning: *M* = 2.21, *SD* = .41; No Warning: *M* = 2.28, *SD* = .43), *t*(947) = 2.77, *p* = .006, *d* = .18, but not novel items, (Warning: *M* = 2.12, *SD* = .38; No Warning: *M* = 2.15, *SD* = .39), *t*(947) = 1.36, *p* = .175, *d* = .09. That is, the warning decreased perceptions of accuracy for items that were presented in the familiarization stage – both fake stories that were labeled with warnings and the real stories presented without warnings[3] – but not for items that were not presented in the familiarization stage.

There was no significant three-way interaction, however, between headline type, exposure, and warning condition, $F < 1$. As a consequence, the repetition effect was evident for fake news headlines in the Warning condition, $t(460) = 4.89$, $p < .001$, $d = .23$, as well as the No Warning condition, $t(487) = 5.81$, $p < .001$, $d = .26$ (see Figure 2). That is, participants rated familiarized fake news headlines that they were explicitly warned about as *more* accurate than novel fake news headlines that they were not warned about (despite the significant negative effect of warnings on perceived accuracy of fake news reported above). In fact, there was no significant interaction between the exposure and warning manipulations when isolating the analysis to fake news headlines, $F(1, 947) = 1.00$, $MSE = .12$, $p = .317$, $\eta^2 = .001$, Thus, the warning seems to have created a general sense of distrust – thereby reducing perceived accuracy for both familiarized and novel fake news headlines – rather than making people particularly distrust the stories that were labeled as disputed.

[3] However, it should be noted that when examining simple effects, there was only a significant negative effect of warning condition on perceived accuracy of familiarized (i.e., warned) fake news (Warning: $M = 1.71$, $SD = .55$; No Warning: $M = 1.82$, $SD = .56$), $t(947) = 3.25$, $p = .001$, $d = .21$, and not familiarized real news (Warning: $M = 2.71$, $SD = .54$; No Warning: $M = 2.74$, $SD = .52$), $t < 1$.

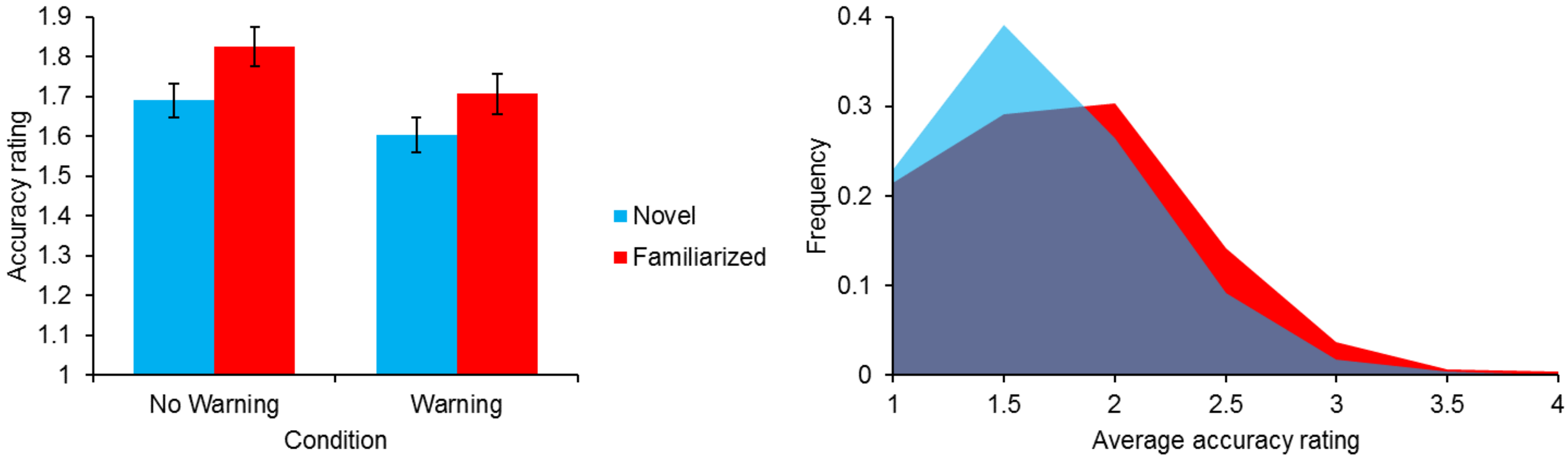


Figure 2. *Exposing participants to fake news headlines in Study 2 increased accuracy ratings, even when the stories were tagged with a warning indicating that they had been disputed by third-party fact checkers. (a) Mean accuracy ratings for fake news headlines as a function of repetition (familiarized stories were shown previously during the familiarization stage; novel stories were shown for the first time during the assessment stage) and presence or absence of a warning that fake news headlines had been disputed. Error bars indicate 95% confidence intervals. (b) Distribution of participant-average accuracy ratings for the fake news headlines, comparing the six familiarized stories shown during the familiarization stage (red) with the six novel stories shown for the first time in the assessment stage (blue). We collapse across warning and no warning conditions as the repetition effect did not differ significantly by condition.*

As a secondary analysis[4], we also investigate whether the effect of prior exposure is robust to political concordance (i.e., whether headlines were congruent or incongruent with one's political stance). Mean perceptions of news accuracy for politically concordant and discordant items as a function of type, exposure, and warning condition can be found in Table 3. Perceived accuracy was entered into a 2 (Political valence: concordant, discordant) x 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA. First, as a manipulation check, politically concordant items were rated as far more accurate than politically discordant items overall (see Table 3), $F(1, 945) = 573.08$, $MSE = .34$, $p < .001$, $\eta^2 = .38$. Nonetheless, we observed no significant interaction between the repetition manipulation and political valence, $F(1, 945) = 2.24$, $MSE = .15$, $p = .135$, $\eta^2 = .002$. The illusory truth effect was evident for fake news headlines that were politically discordant, $t(946) = 4.70$, $p < .001$, $d = .15$, as well as concordant, $t(946) = 7.19$, $p < .001$, $d = .23$. Political concordance interacted significantly with type of news story, $F(1, 945) = 138.91$, $MSE = .23$, $p < .001$, $\eta^2 = .13$, such that the difference between perceptions of real and fake news (i.e., discernment) was greater for politically concordant headlines (Real: $M = 2.90$, $SD = .59$; Fake: $M = 1.80$, $SD = .56$), than politically discordant headlines (Real: $M = 2.44$, $SD = .53$; Fake: $M = 1.61$, $SD = .48$), $t(946) = 11.8$, $p < .001$, $d = .38$ (see Pennycook & Rand, 2017a for a similar result). All other interactions with political concordance were not significant, all F's < 1.5, p's > .225.

[4] These analyses were not preregistered, although we did preregister a parallel analysis where pro-Democrat and pro-Republican items would be analyzed separately while comparing liberals and conservatives. The present analysis simply combines the data into a more straightforward analysis and uses the binary Clinton/Trump choice to distinguish liberals and conservatives.

Table 3. Means, standard deviations, and significance tests (comparing familiarized and novel items) for politically concordant and discordant items in the warning and no warning conditions. Political concordant items consistent of Pro-Democrat items for Clinton supporters and Pro-Republican items for Trump supporters (and vice-versa for politically discordant items).

	Type	Warning	Familiarized	Novel	*t* (df)	*p*
Politically Concordant	Fake News	No Warning	1.93 (0.7)	1.78 (0.6)	5.46 (486)	< .001
		Warning	1.81 (0.7)	1.68 (0.6)	4.69 (459)	< .001
	Real News	No Warning	2.98 (0.6)	2.83 (0.7)	5.45 (486)	< .001
		Warning	2.92 (0.7)	2.86 (0.7)	2.17 (459)	.031
Politically Discordant	Fake News	No Warning	1.72 (0.6)	1.60 (0.5)	3.91 (486)	< .001
		Warning	1.60 (0.6)	1.53 (0.5)	2.66 (459)	.008
	Real News	No Warning	2.50 (0.6)	2.39 (0.6)	3.85 (486)	< .001
		Warning	2.49 (0.6)	2.40 (0.6)	3.03 (459)	.003

The illusory truth effect also persisted when analyzing only news headlines that the participants marked as unfamiliar (i.e., in the same mixed ANOVA as above but only analyzing stories the participants were *not* consciously aware of having seen in the familiarization stage or at some point prior to the experiment) (Familiarized: $M = 1.90$, $SD = .53$; Novel: $M = 1.83$, $SD = .49$), $F(1,541)$[5] $= 11.82$, $MSE = .17$, $p = .001$, $\eta^2 = .02$. See SI for details and further statistical analysis.

Discussion

The results of Study 2 indicate that a single prior exposure is sufficient to increase perceived accuracy for both fake and real news. This occurs even 1) when fake news is labeled as "Disputed by 3rd party fact-checkers" during the familiarization stage (i.e., during encoding at first exposure), 2) among fake (and real) news headlines that are *inconsistent* with one's political ideology, and 3) when isolating the analysis to news headlines that participants were not consciously aware of having seen in the familiarization stage.

[5] Degrees of freedom are lower here because this analysis only includes individuals who were unfamiliar with at least one item in each cell of the design (familiarized/novel and fake/real).

Study 3 – Fake News, One Week Interval

We next sought to assess the robustness of our finding that repetition increases perceptions of fake news accuracy by making two important changes to the design of Study 2. First, we assessed the persistence of the repetition effect by inviting participants back after a week long delay (following previous research which has shown illusory truth effects to persist over substantial periods of time, e.g. Hasher, Goldstein, & Toppino, 1977; Schwartz, 1982). Second, we restricted our analyses to only those items that were unfamiliar to participants when entering the study, which allows for a cleaner 'novel' baseline.

Method

Participants. Our target sample was 1000 participants from Mechanical Turk. This study was completed on February 1st and 2nd, 2017. Participants who completed Study 1 were not permitted to complete Study 2. In total, 1032 participants completed the study, 40 of which dropped out or had missing data (14 from the no warning condition, 27 from the warning condition). Participants who reported responding randomly (N = 29), skipping over the familiarization phase (N = 1), or searching online for the headlines (N = 22) were removed. These exclusions were preregistered. The final sample (N = 940; Mean age = 36.8) included 436 males and 499 females (5 did not respond).

Materials and Procedure. The design was identical to Study 2 (including the Warning and No Warning conditions), with a few exceptions. First, the length of the distractor stage was increased by adding 20 unrelated questionnaire items to the demographics questions (namely, the PANAS, as in Study 1). This filler stage took approximately two minutes to complete. Furthermore, participants were invited to return for a follow-up session one week later in which they were presented with the same headlines they had seen in the assessment stage plus a set of novel headlines not included in the first session (N = 566 participants responded to the follow-up invitation). To allow full counterbalancing, we presented participants with 8 headlines in the familiarization phase, 16 headlines in the accuracy judgment phase (of which 8 were those shown in the familiarization phase), and 24 headlines in the follow-up session a week later (of which 16 were those shown in the assessment phase of the first session), again maintaining an equal number of real/fake and Pro-Democrat/Pro-Republican headlines within each block. The design of Study 2 therefore allowed us to assess the temporal stability of the repetition effect both within session 1 (over the span of a distractor task) and session 2 (over the span of a week).

Second, during the familiarization stage participants were asked to indicate whether each headline was familiar, instead of whether they would share the story on social media (the social media question was moved to the assessment stage). This modification allowed us to restrict our analyses to only those items that were unfamiliar to participants when entering the study (i.e., they said "no" when asked about familiarity)[6], allowing for a cleaner assessment of the causal effect of repetition (903 of the 940 participants in the first session were previously unfamiliar with at least one story of each type and thus included in the main text analysis, as were as were 527 out of the 566 participants in second session; see SI for analyses with all items and all participants).

[6] Whereas participants indicated their familiarity with familiarized items prior to completing the accuracy judgments, they indicated their familiarity with novel items after completing the accuracy judgments. Thus, it is possible that seeing the news headlines in the accuracy judgment phase would increase perceived familiarity. There was no evidence for this, however, as mean familiarity judgment (scored continuously) did not significantly differ based on whether the judgment was made before or after the accuracy judgment phase, *t*(939) = 0.68, *SE* = .01, *p* = .494, *d* = .02. Participants were unfamiliar with 81.2% of the fake news headlines and 49.2% of the real news headlines.

As in Experiment 2, our preregistration specified the comparison between familiarized and novel fake news in both warning and no warning conditions (and for both sessions) as the key analyses, although in this case we preregistered the full 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA. We also preregistered the political concordance analysis. Finally, we preregistered the removal of cases where participants were familiar with the news headlines as a secondary analysis, but we will focus on it as a primary analysis here as this is the novel feature relative to Study 2 (primary analyses including all participants are discussed in footnote 8). Our preregistration is available at the following link: https://osf.io/txf46/.

Results

Perceived accuracy was entered into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA. Replicating the illusory truth effect from Study 2, there was a clear causal effect of prior exposure on accuracy in the first session of Study 3 despite the longer distractor stage: Headlines presented in the familiarization stage (*M* = 2.01, *SD* = .54) were rated as more accurate than novel headlines (*M* = 1.92, *SD* = .49), *F*(1, 721) = 22.52, *MSE* = .23, $p < .001$, $\eta^2 = .03$. Again replicating the results of Study 2, while there was a significant main effect of type, such that real stories (*M* = 2.31, *SD* = .63) were rated as much more accurate than fake stories (*M* = 1.63, *SD* = .52), *F*(1, 721) = 934.57, *MSE* = .36, $p < .001$, $\eta^2 = .57$, there was no significant interaction between exposure and type of news headline, *F*(1, 721) = 2.65, *MSE* = .20, *p* = .104, $\eta^2 = .004$. Accordingly, prior exposure increased perceived accuracy even when only considering fake news headlines (see Figure 3a,b), *t*(902)[7] =

[7] Only unfamiliar headlines are included and therefore missing data accounts for missing participants in some cell of the design. Degrees of freedom vary throughout because the maximum number of participants is included in each analysis.

5.99, $p < .001$, $d = .20$ (89.5% increase in number of participants judging familiarized fake news headlines as accurate compared to novel fake news headlines, from 38 to 72 participants out of 903).

Unlike Study 2, there was no main effect of the warning manipulation, $F < 1$. However, there was a marginally significant interaction between type of news story and warning condition, $F(1, 721) = 2.95$, $MSE = .36$, $p = .086$, $\eta^2 = .004$. Regardless, the fake news warnings in the familiarization stage had no significant overall effect on perceptions of fake news accuracy in the assessment stage (Warning: $M = 1.61$, $SD = .50$; No Warning: $M = 1.66$, $SD = .54$), $t(932)$[8] = 1.54, $p = .123$, $d = .10$. There was also no significant effect of the warning on perceptions of real news accuracy (Warning: $M = 2.32$, $SD = .63$; No Warning: $M = 2.30$, $SD = .63$), $t < 1$, no significant interaction between the repetition and warning manipulations, $F(1,721) = 1.89$, $MSE = .23$, $p = .169$, $\eta^2 = .003$), and no significant three-way interaction between warning, exposure, and type of news story, $F < 1$.[9] Nonetheless, it should be noted that familiarized fake news headlines (i.e., the fake news headlines that were warned about in the familiarization stage) were rated as less accurate ($M = 1.64$, $SD = .59$) than the same headlines in the control (no warning) condition ($M = 1.73$, $SD = .63$), $t(925) = 2.14$, $p = .032$, $d = .14$, suggesting that the warning did have some effect on accuracy judgments.

[8] Degrees of freedom change here because this analysis includes the maximum number of individuals who were unfamiliar with at least one fake news item.

[9] In our (also) preregistered analysis that includes both previously familiar and unfamiliar items, there is a main effect of repetition, $F(1, 938) = 18.98$, $MSE = .16$, $p < .001$, $\eta^2 = .02$, but (unlike in Study 2) a significant interaction between exposure and warning condition, $F(1, 938) = 7.81$, $MSE = .16$, $p = .005$, $\eta^2 = .01$. There was a significant repetition effect for fake news in the no warning condition, $t(475) = 5.31$, $SE = .03$, $p < .001$, $d = .24$, but no effect in the warning condition, $t(463) = 1.30$, $SE = .03$, $p = .193$, $d = .06$. It is possible that prior knowledge of the items facilitated explicit recall of the warning, which may have mitigated the illusory truth effect. See SI for means and further analyses.

Following our preregistration, we also analyzed the effect of exposure for fake news headlines separately in the warning and no warning conditions. The repetition effect was evident for fake news headlines in both the Warning condition (Familiarized: *M* = 1.63, *SD* = .58; Novel: *M* = 1.55, *SD* = .52), *t*(447) = 3.07, *p* = .002, *d* = .14, and the No Warning condition (Familiarized: *M* = 1.71, *SD* = .61; Novel: *M* = 1.58, *SD* = .54), *t*(454) = 5.41, *p* < .001, *d* = .25. Furthermore, familiarized fake news headlines were judged as more accurate than novel ones for both political discordant (Familiarized: *M* = 1.60, *SD* = .67; Novel: *M* = 1.51, *SD* = .63), *t*(858) = 3.41, *p* = .001, *d* = .12, and concordant items (Familiarized: *M* = 1.72, *SD* = .77; Novel: *M* = 1.59, *SD* = .67), *t*(801) = 4.93, *p* < .001, *d* = .18 (an ANOVA including concordance indicated that there was no significant interaction between repetition and political concordance for fake news, *F*(1,769) = 1.46, *MSE* = .32, *p* = .228, η^2 = .002)[10].

Following up one week later, we continued to find a clear causal effect of repetition on accuracy ratings: Perceived accuracy of a story increased linearly with the number of times the participants had been exposed to that story. Using linear regression with robust standard errors clustered on participant[11], we found a significant positive relationship between number of exposures and accuracy overall, *b* = .043, *t*(526) = 3.48, *p* = .001, and when only considering

[10] We focus on fake news headlines here because the political concordance manipulation cuts the number of items in half. Including real news in this analysis decreases the number of participants markedly because the ANOVA requires each participant to contribute at least one observation to each cell of the design. Nonetheless, the full ANOVA reveals a significant main effect of repetition, *F*(1,312) = 8.94, *p* = .003, η^2 = .03, and no interaction with political concordance, *F* < 1.

[11] This specific analysis was not preregistered. Rather, the preregistration called for a comparison of the full 16 items from session 1 with the 8 novel items in session 2. This, too, revealed a significant main effect of repetition (using the same ANOVA as in the session 1 analysis), *F*(1,453) = 12.91, *p* < .001, η^2 = .03. However, such an analysis does not tell us about the increasing effect of exposure, hence our deviation from the preregistration. See SI for further details and analyses.

fake news headlines (see Figure 3c), *b* = .048, *t*(526) = 3.66, *p* < .001 (64% increase in number of participants judging fake news headlines as accurate among stories seen twice compared to novel fake news headlines, from 25 to 41 participants out of 527). Once again, this relationship was evident for fake news in both the Warning condition, *b* = .036, *t*(276) = 1.97, *p* = .050, and the No Warning condition, *b* = .061, *t*(249) = 3.27, *p* = .001 (there was no significant interaction between the repetition and warning manipulations, *b* = -.025, *t*(526) = 0.96, *p* = .337; see Figure 3c); and for fake news headlines that were politically discordant, *b* = .041, *t*(525) = 2.28, *p* = .023, as well as concordant, *b* = .061, *t*(523) = 3.24, *p* = .001.

Discussion

The results of Study 3 further demonstrated that prior exposure increases perceived accuracy of fake news. This occurred regardless of political discordance and among previously unfamiliar headlines that were explicitly warned about during familiarization. Crucially, the effect of repetition on perceived accuracy persisted after a week and increased with an additional repetition. This suggests that fake news credulity compounds with increasing exposures and maintains over time.

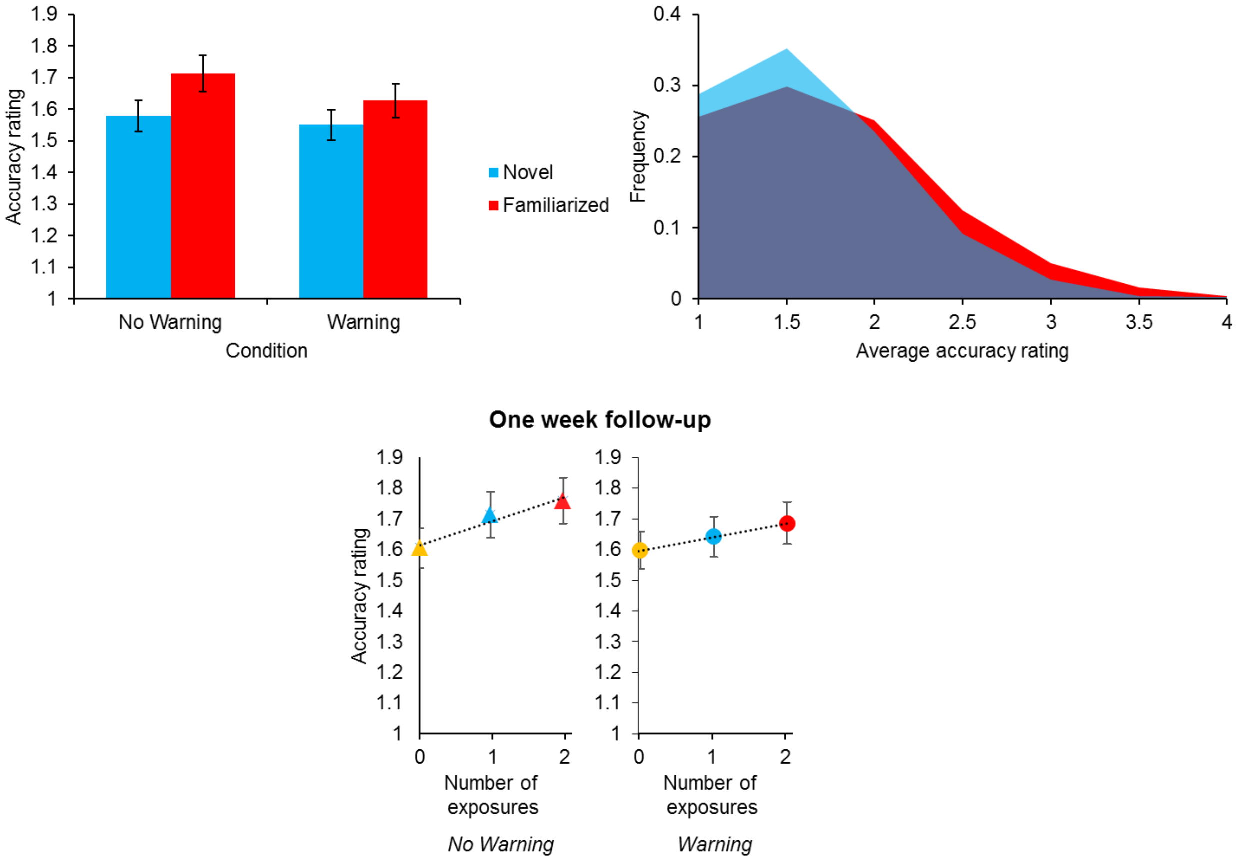


Figure 3. *The illusory truth effect for fake news is persistent, lasting over a longer filler stage in Study 2 and continuing to be observed in a follow-up session one week later. (a) Mean accuracy ratings for fake news headlines in the initial session of Study 2 as a function of repetition and presence or absence of a warning that fake news headlines had been disputed. Error bars indicate 95% confidence intervals. (b) Distribution of participant-average accuracy ratings for the fake news headlines in Study 2, comparing the four headlines shown during the familiarization stage (red) with the four novel headlines shown for the first time in the assessment stage (blue). We collapse across warning and no warning conditions as the repetition effect did not differ significantly by condition. (c) Mean accuracy ratings for fake news headlines in the follow-up session conducted one week later, as a function of number of exposures to the story (2 times for headlines previously presented in the familiarization and assessment stage of the first session; 1 time for headlines previously presented only in the assessment stage of the first session; and 0 times for headlines introduced for the first time in the follow-up session) and presence or absence of warning tag. Error bars indicate 95% confidence intervals based on robust standard errors clustered by participant, and trend line shown in dotted black.*

Study 4 – Item-Level Analysis of Plausibility

Taken together, the results of Study 1, where extremely implausible and plausible trivia statements did *not* show a significant illusory truth effect, and Studies 2 and 3, where fake news headlines that were highly implausible (but not as extreme as the trivia in Study 1) *did* show a significant illusory truth effect, have important theoretically implications regarding the impact of plausibility on illusory truth. Study 4 uses the data from Studies 1-3 to shed further light on this issue.

While prior theory has not explicitly considered the impact of plausibility, the Fluency-Conditional Model proposed by Fazio et al. (2015) provides a framework in which to ground our inquiry. According to this model, the first step in accuracy judgments is the determination of whether a participant will rely on fluency, or will not do so either because fluency is absent (e.g. the statement is novel) or because fluency is present (e.g. the statement has been familiarized) but gets discounted by the participant. It is only in the cases where people do not rely on fluency that prior knowledge regarding veracity enters the judgment process. The size of the illusory truth effect is thus determined by the extent to which experimental exposure increases fluency, and is independent of knowledge (except for potential ceiling effects) – as is consistent with the data reported by Fazio and colleagues.

The lack of illusory truth effect for the extremely implausible statements in our Study 1, however, suggests that prior knowledge *can* actually exert a greater influence earlier in the process than suggested by the Fluency-Conditional Model, and thus influence the size of the illusory truth effect. That is, it seems that very salient knowledge-based information about veracity can increase the probability that fluency is discounted, so much so that no repetition effect appears whatsoever. Nonetheless, the question remains: Is this disruption of fluency by knowledge unique to these boundary cases, such that knowledge only influences fluency when it is completely obvious that statements are false? Or is there a more continuous relationship, such that knowledge plays an important role in whether people rely on fluency more generally? Put differently, although the Fluency-Conditional Model proposes a serial processing chain that moves from fluency to knowledge, it is possibly that these components *interact* to determine judgments of accuracy. This is consistent with recent dual-process models of reasoning and decision making wherein multiple intuitive outputs based on low-level autonomous responses (e.g., based on automatic memory retrieval or fluency heuristics) interact to initiate high-level reasoning processes (which influence perceptions of accuracy) (De Neys, 2012; Pennycook, Fugelsang, & Koehler, 2015). To investigate this issue quantitatively, we pool the data from Studies 1 through 3, and conduct an item-level analysis of the relationship between baseline plausibility and the size of the illusory truth effect.

Method

Across our three studies, a total of 76 different statements were used. For each of the 76 statements, we define baseline plausibility as the average accuracy rating (normalized to the interval [0,1]) when novel (i.e. not familiarized), and the illusory truth effect size as the average accuracy rating when familiarized minus when novel (both normalized to the interval [0,1]).[12] As we are focusing here on plausibility, we do not include the Warning conditions from Studies 2 and 3; and because excluding previously familiar headlines in Study 3 may cause selection effects for real news (since many people were familiar with many of these headlines), we do not filter based on familiarity. However, our results do not depend on these exclusions.

[12] At the item level, these differences are between-subject since each item was either *only* novel or *only* familiarized in the assessment stage for an individual participant.

Based on the lack of illusory truth effect for the extremely implausible items in Study 1, but presence of effect for the fake news headlines in Studies 2 and 3, we predict that the effect size will be increasing with baseline plausibility when plausibility is low. By necessity, however, the effect size must eventually begin to decline when plausibility becomes sufficiently high because of ceiling effects.[13] Thus, there is reason to expect a curvilinear relationship between baseline plausibility and the size of the illusory truth effect. This analysis was conducted following the completion of our prior 3 studies, and was not itself preregistered.

Results

As shown in Figure 4, we do indeed observe such a curvilinear relationship across the 76 statements tested in Studies 1-3. A regression predicting illusory truth effect size shows a significant positive linear effect of baseline plausibility, $\beta = 1.063$, $t(73) = 3.21$, $p = .002$, and a significant negative quadratic effect of baseline plausibility, $\beta = -1.259$, $t(73) = -3.81$, $p < .001$. When including indicator dummy variables for veracity (0 = false/fake, 1 = true/real) and statement type (0 = trivia, 1 = news headline), we continue to find a significant positive linear effect of baseline plausibility, $\beta = 1.356$, $t(71) = 3.91$, $p < .001$, and a significant negative quadratic effect of baseline plausibility, $\beta = -1.565$, $t(71) = -4.74$, $p < .001$, as well as a significant negative effect of statement type such that news headlines showed smaller effects, $\beta = -.306$, $t(71) = -2.70$, $p = .009$. There was no significant effect of veracity, $\beta = -.113$, $t(71) = -0.74$, $p = .464$. The inverted U-shaped nature of the relationship is further demonstrated by separate linear regressions which show a significant positive relationship when baseline plausibility is

[13] Note that floor effects cannot explain the positive association between increasing plausibility and the illusory truth effect at the bottom end of the scale because the effect is moving away from the floor.

below the midpoint of the scale (0.5), *β* = .337, *t*(41) = 2.29, *p* = .027, and a significant negative relationship when baseline plausibility is above the midpoint, *β* = -.452, *t*(31) = -2.82, *p* = .008.

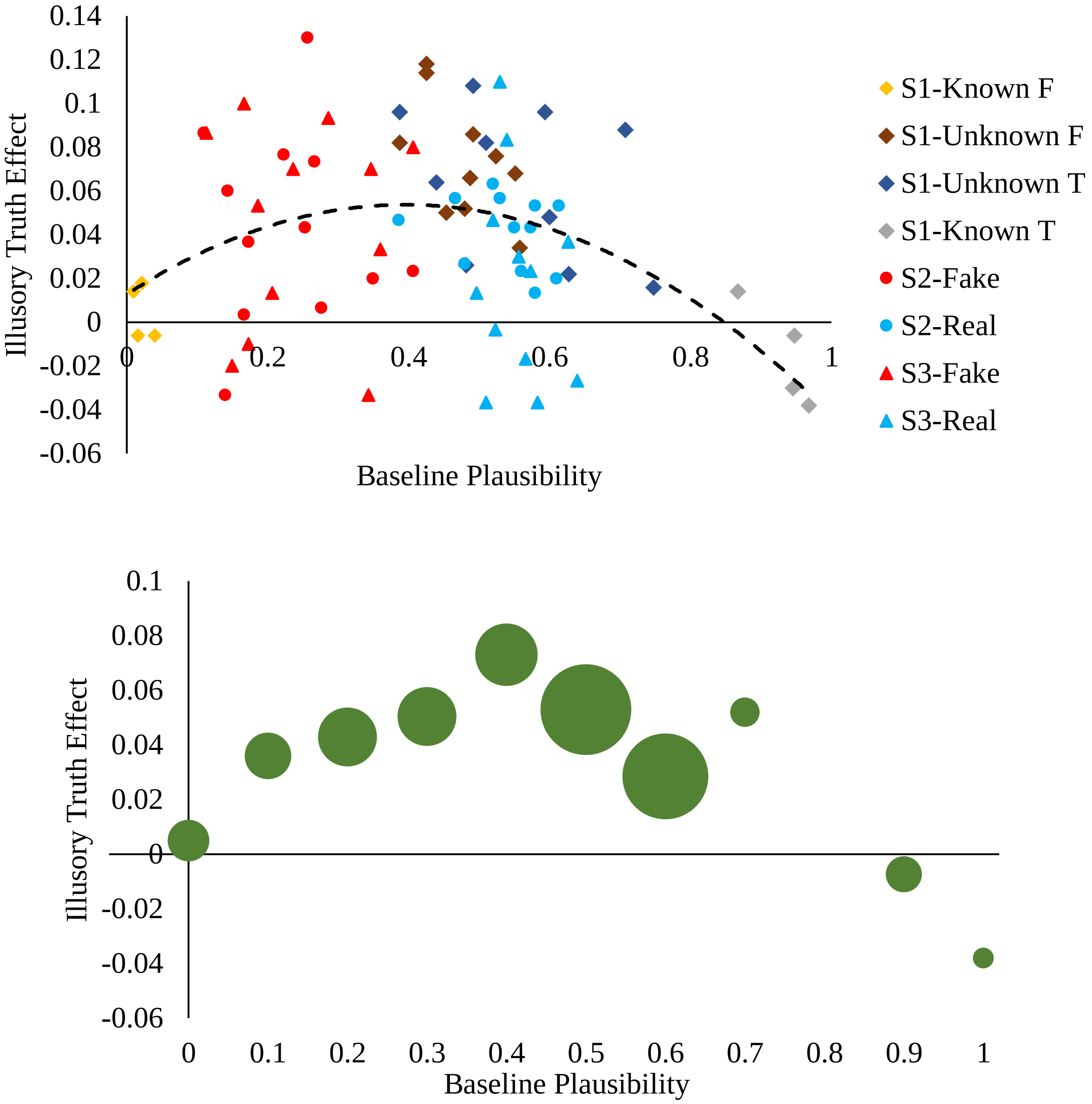


Figure 4. A curvilinear relationship exists between a statement's baseline plausibility (mean accuracy in the novel condition) and the size of the illusory truth effect for that statement. (a) Each data point represents one statement for either Study 1 (diamonds), Study 2 (circles) or Study 3 (triangles). The regression line with linear and quadratic terms is shown. (b) Baseline plausibility scores binned by rounding to the nearest 0.1, and the mean illusory truth effect size for each bin is shown; size of dots is proportional to the number of statements in each bin.

Discussion

These results demonstrate that plausibility plays an important role in the illusory truth effect across the full range of plausibility levels, not just at the extremes. Thus, they necessitate a revision to the Fluency-Conditional Model of illusory truth in which prior knowledge not only affects judgments when fluency is absent or discounted, but also influences the likelihood that participants rely on fluency in the first place.

General Discussion

While repetition did not impact accuracy judgments of totally implausible statements, across two preregistered experiments with a total of more than 1,800 participants we found consistent evidence that repetition *did* increase the perceived accuracy of fake news headlines. Indeed, a single prior exposure to fake news headlines was sufficient to measurably increase subsequent perceptions of their accuracy. Although this effect was relatively small (d = .20-.21), it increased with a second exposure, thereby suggesting a compounding effect of repetition across time. Explicitly warning individuals that the fake news headlines had been disputed by third-party fact-checkers (which was true in every case) did not abolish or even significantly diminish this effect. The illusory truth effect was also evident even among news headlines that were inconsistent with the participants' stated political ideology. Finally, a combined analysis of our three experiments revealed a quadratic association between baseline plausibility and the size of the illusory truth effect, such that there is no effect for extremely implausible content, but it emerges and increases in size as statements become more plausible at the bottom end of the scale. The illusory truth effect then begins to decrease as plausibility moves beyond the scale midpoint and towards the ceiling of the scale.

Mechanisms of illusory truth

First, it is important to note that repetition increased accuracy even for items that the participants were not consciously aware of having been exposed to. This supports the broad consensus that repetition influences accuracy through a low-level fluency heuristic (Alter & Oppenheimer, 2009; Begg et al., 1992; Reber et al., 1998; Unkelbach, 2007; Whittlesea, 1993). These findings indicate that our repetition effect is likely driven, at least in part, by automatic (as opposed to strategic) memory retrieval (Diana, Yonelinas, & Ranganath, 2007; Yonelinas, 2002; Yonelinas & Jacoby, 2012). More broadly, these effects correspond with prior work demonstrating the power of fluency to influence a variety of judgments (Schwarz, Sanna, Skurnik, & Yoon, 2007) – for example, subliminal exposure to a variety of stimuli (e.g., Chinese characters) increases associated positive feelings (i.e., the mere exposure effect; see Zajonc, 1968, 2001). Our evidence that the illusory truth effect extends to implausible and even politically inconsistent fake news stories expands the scope of these effects. That perceptions of fake news accuracy can be manipulated so easily despite being highly implausible (only 15-22% of the headlines were judged to be accurate) has substantial practical implications (discussed below) – but what implications do these results have for our understanding of the mechanisms that underlie the illusory truth effect (and, potentially, a broader array of fluency effects observed in the literature)?

For decades, it had been assumed that repetition only increases accuracy in cases where the individual does not have prior knowledge. However, recent evidence indicates that repetition can even increase the perceived accuracy of plausible but false statements (e.g. “chemosynthesis is the name of the process by which plants make their food”) among participants who were subsequently able to identify the correct answer (Fazio et al., 2015). The Fluency-Conditional Model was proposed to explain this effect, whereby participants only stop to consult their prior knowledge for statements that are not fluent. As shown in Study 1, however, repetition does *not* increase perceived accuracy for extremely implausible claims (e.g., “The Earth is a perfect square”), indicating that prior knowledge *is* deployed even for fluent statements when it is sufficiently obvious that statements are false. Moreover, as is evident from Study 4, plausibility has an impact on the size of the illusory truth effect across the full range of baseline plausibility ratings, such that the effect increases as statements are judged to be increasingly likely to be true (and this increase continues until ceiling effects lead to a reversal). These results indicate that, contrary to the Fluency-Conditional Model, prior knowledge and fluency *interact* to determine judgments of accuracy. What these results suggest, then, is a model of the illusory truth effect in which inputs based on knowledge and fluency are generated in parallel. This is in-keeping with recent accounts of high-level reasoning where parallel low-level processes have been used to explain the initiation of deliberation during judgment and decision-making (De Neys, 2012; Pennycook, Fugelsang, & Koehler, 2015). Furthermore, researchers have distinguished between *answer* fluency (how quickly and easily a response comes to mind) and *processing* fluency (how quickly and easily a stimulus is processed) (Thompson et al., 2013a; Thompson et al., 2013b). One possibility is that repetition increases processing fluency (Unkelbach, 2006) whereas prior knowledge increases answer fluency, and these two factors compete to determine perceptions of accuracy. Illuminating how exactly knowledge and fluency interact to produce the inverted U-shaped relationship we observe is an important direction for future research.

Motivated reasoning

Our results also have implications for a broad debate about the scope of motivated reasoning, which has been taken to be a fundamental aspect of how individuals interact with political misinformation and disinformation (Swire, Berinsky, Lewandowsky, & Ecker, 2017) and has been used to explain the spread of fake news (Allcott & Gentzkow, 2017; Beck, 2017; Calvert, 2017; Kahan, 2017; Singal, 2017). While Trump supporters were indeed more skeptical about fake news headlines that were Anti-Trump relative to Clinton supporters (and vice versa), remarkably our results show that repetition increases perceptions of accuracy even in such politically discordant cases. Take, for example, the item “BLM Thug Protests President Trump with Selfie… Accidentally Shoots Himself In The Face,” which is politically discordant for Clinton supporters and politically concordant for Trump supporters. While on first exposure Clinton supporters were less likely (11.7%) to rate this headline as accurate than Trump supporters (18.5%), suggesting the potential for motivated reasoning, a single prior exposure to this headline increased accuracy judgements in both cases (to 17.9% and 35.5%, for Clinton and Trump supporters respectively). Thus, fake news headlines were positively affected by repetition even when there was a strong political motivation to reject them. This observation complements the results of Pennycook and Rand (2017a), who find – in contrast common motivated reasoning accounts (Kahan, 2017) – that analytic thinking leads to disbelief in fake news regardless of political concordance. Taken together, this suggests that motivated reasoning may actually play less of a role in the spread of fake news than is often argued.

These results also bear on a recent debate about whether corrections might actually make false information more familiar, thereby increasing the incidence of subsequent false beliefs (i.e., the familiarity backfire effect; Berinsky, 2015; Nyhan & Reifler, 2010; Schwarz et al., 2007; Skurnik et al., 2005). In contrast to the backfire account, the latest research in this domain indicates that explicit warnings or corrections of false statements actually have a positive (and certainly not negative) impact on subsequent perceptions of accuracy (Ecker, Hogan, & Lewandowsky, 2017; Lewandowsky et al., 2012; Pennycook & Rand, 2017b; Swire, Ecker, et al., 2017). In our data, the positive effect of a single prior exposure (*d* = .20 in Study 2) was effectively equivalent to the negative effect of the "Disputed" warning (*d* = .17 in Study 2). Thus, although any benefit arising from the disputed tag is immediately wiped out by the prior exposure effect, we also do not find any evidence of a meaningful backfire. Our findings therefore support recent skepticism about the robustness and importance of the familiarity backfire effect.

Societal implications

Our findings have important implications for the functioning of democracy, which relies on an informed electorate. Specifically, our results shed some light on what can be done to combat belief in fake news. We employed a warning that was developed by Facebook to curb the influence of fake news on their social media platform ("Disputed by 3rd Party Fact-Checkers"). We found that this warning did not disrupt the illusory truth effect, an observation that resonates with previous work demonstrating that, for example, explicitly labelling consumer claims as false (Skurnik et al., 2005) or retracting pieces of misinformation in news articles (Berinsky, 2015; Ecker, Lewandowsky, & Tang, 2010; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012; Nyhan & Reifler, 2010) are not necessarily effective strategies for decreasing long-term misperceptions (but see Swire, Ecker, & Lewandowsky, 2017). Nonetheless, it is important to note that the warning *did* successfully decrease subsequent overall perceptions of the accuracy of fake news headlines; the warning's effect was just not specific to the particular fake news headlines that the warning was attached to (and so the illusory truth effect survived the warning). Thus, the warning appears to have increased general skepticism, which increased the overall sensitivity to fake news (i.e., the warning decreased perceptions of fake news accuracy without affecting judgments for real news). The warning also successfully decreased people's willingness to share fake news headlines on social media. However, neither of these warning effect sizes were particularly large – for example, as described above, the negative impact of the warning on accuracy was entirely canceled out by the positive impact of repetition. That result, coupled with the persistence of the illusory truth effect we observed and the possibility of an "implied truth" effect whereby tagging some fake headlines may increase the perceived accuracy of *untagged* fake headlines (Pennycook & Rand, 2017b), suggests that larger solutions are needed that prevent people from ever seeing fake news in the first place, rather than qualifiers aimed at making people discount the fake news that they do see.

Finally, our findings have implications beyond just fake news on social media. They suggest that politicians who continuously repeat false statements will be successful, at least to some extent, in convincing people those statements are in fact true. Indeed, the word "delusion" derives from a Latin term conveying the notion of mocking, defrauding, and deception. And the illusory truth effect for highly salient and impactful information we demonstrate here suggests that familiarity may also play an important role in domains beyond politics, such as the formation of religious and paranormal beliefs where claims are difficult to either validate or reject empirically. When the truth is hard to come by, familiarity is an attractive stand-in.

Context

In this research program, we use cognitive psychological theory and techniques to illuminate issues that have clear consequences for everyday life, with the hope of generating insights that are both practically and theoretically relevant. The topic of fake news – and disinformation more broadly – is of great relevance to current public discourse and policy making, and fits squarely in the domain of cognitive psychology. Plainly, this topic is something that cognitive psychologists should be able to say something specific and illuminating about!

Author Note. This research was supported by a Social Sciences and Humanities Council of Canada Banting Postdoctoral Fellowship (to G. P.), a grant from the National Institute of Mental Health (MH081902) (to T.D.C.), and grants from the Templeton World Charity Foundation and DARPA (to D.G.R.). A working paper version of the manuscript was posted online via the Social Sciences Research Network (https://ssrn.com/abstract=2958246) and on ResearchGate (https://www.researchgate.net/publication/317069544_Prior_Exposure_Increases_Perceived_Accuracy_of_Fake_News). Portions of this research were presented in 2017 at the following venues: 1) Conference for Combating Fake News: An Agenda for Research and Action. Harvard Law School & Northeastern University, 2) Canadian Society for Brain, Behaviour and Cognitive Science Annual Conference, 3) Brown University, Department of Psychology, 4) Harvard University, Department of Psychology, 5) Yale University, Institute for Network Science and Technology and Ethics Study Group, and 6) University of Connecticut, Cognitive Science Colloquium.

References

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *NBER Working Paper No. 23098*. Retrieved from http://www.nber.org/papers/w23089

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

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*(6), 576–605. https://doi.org/10.1016/0022-1031(91)90026-3

Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making*, *2*(2), 81–94. https://doi.org/10.1002/bdm.3960020203

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning & Memory*, *5*(3), 241–252. https://doi.org/10.1037/0278-7393.5.3.241

Beck, J. (2017). This article won’t change your mind: The fact on why facts alone can’t fight false beliefs. Retrieved August 2, 2017, from https://www.theatlantic.com/science/archive/2017/03/this-article-wont-change-your-mind/519093/

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446–458. https://doi.org/10.1037/0096-3445.121.4.446

Berinsky, A. A. J. (2015). Rumors and Health Care Reform: Experiments in Political Misinformation. *British Journal of Political Science*, (June), 1–22.

https://doi.org/10.1017/S0007123415000186

Calvert, D. (2017). The Psychology Behind Fake News. Retrieved August 2, 2017, from https://insight.kellogg.northwestern.edu/article/the-psychology-behind-fake-news

Coppock, A. (2016). Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach. Retrieved from https://alexandercoppock.files.wordpress.com/2016/02/coppock_generalizability2.pdf

Corlett, P. R. (2009). Why do delusions persist? *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.012.2009

De Neys, W. (2012). Bias and Conflict: A Case for Logical Intuitions. *Perspectives on Psychological Science*, *7*(1), 28–38. https://doi.org/10.1177/1745691611429354

Dechene, A., Stahl, C., Hansen, J., & Wanke, M. (2010). The Truth About the Truth: A Meta-Analytic Review of the Truth Effect. *Personality and Social Psychology Review*, *14*(2), 238–257. https://doi.org/10.1177/1088868309352251

Diana, R., Yonelinas, A., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1364661307001878

Ecker, U., Hogan, J., & Lewandowsky, S. (2017). Reminders and Repetition of Misinformation: Helping or Hindering Its Retraction? *Journal of Applied Research in Memory and Cognition*, *6*, 185–192.

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. https://doi.org/10.3758/MC.38.8.1087

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/http://dx.doi.org/10.1037/xge0000098

Flynn, D., Nyhan, B., & Reifler, J. (2016). The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs about Politics. *Advances in Pol. Psych*. Retrieved from http://djflynn.org/wp-content/uploads/2016/11/FlynnNyhanReifler.pdf

Gabielkov, M., Ramachandran, A., & Chaintreau, A. (2016). Social Clicks: What and Who Gets Read on Twitter? *ACM SIGMETRICS 2016*. Retrieved from http://dl.acm.org/citation.cfm?id=2901462

Gigerenzer, G. (1984). External Validity of Laboratory Experiments: The Frequency-Validity Relationship. *The American Journal of Psychology*, *97*(2), 185. https://doi.org/10.2307/1422594

Goldman, R. (2016). Reading Fake News, Pakistani Minister Directs Nuclear Threat at Israel. Retrieved March 2, 2017, from https://www.nytimes.com/2016/12/24/world/asia/pakistan-israel-khawaja-asif-fake-news-nuclear.html

Gottfried, J., & Shearer, E. (2016). News Use Across Social Media Platforms 2016. Retrieved March 2, 2017, from http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

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. https://doi.org/10.1016/S0022-5371(77)80012-1

Hawkins, S., & Hoch, S. (1992). Low-involvement learning: Memory without evaluation. *Journal of Consumer Research*. Retrieved from http://jcr.oxfordjournals.org/content/19/2/212.abstract

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*. Retrieved from http://www.springerlink.com/index/8V022H46LP756774.pdf

Johar, G., & Roggeveen, A. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology*. Retrieved from http://www.sciencedirect.com/science/article/pii/S1057740807700189

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424. https://doi.org/10.2139/ssrn.2182588

Kahan, D. M. (2017). Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2973067

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732–735. https://doi.org/10.1038/nclimate1547

Krupnikov, Y., & Levine, A. (2014). Cross-sample comparisons and external validity. *Journal of Experimental Political Science*, *1*, 59–80. Retrieved from https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/cross-sample-comparisons-and-external-validity/B11437F96788A7F01653A7C1C9E87F34

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

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

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *The Behavioral and Brain Sciences*, *34*(2), 57-74-111. https://doi.org/10.1017/S0140525X10000968

Mosseri, A. (2016). Building a Better News Feed for You. Retrieved March 2, 2017, from http://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you/

Mullinix, K., Leeper, T., Druckman, J., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*, 109–138. Retrieved from https://www.cambridge.org/core/journals/journal-of-experimental-political-science/article/the-generalizability-of-survey-experiments/72D4E3DB90569AD7F2D469E9DF3A94CB

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

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2015). What makes us think? A three-stage dual-process model of analytic engagement. *Cognitive Psychology*, *80*, 34–72. https://doi.org/10.1016/j.cogpsych.2015.05.001

Pennycook, G., & Rand, D. G. (2017, August 21). Who Falls for Fake News? The Roles of Analytic Thinking, Motivated Reasoning, Political Ideology, and Bullshit Receptivity.

Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3023545

Polage, D. C. (2012). Making up history: False memories of fake news stories. *Europe's Journal of Psychology*, *8*(2), 245–250. https://doi.org/10.5964/ejop.v8i2.456

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48.

Redlawsk, D. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *Journal of Politics*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/1468-2508.00161/full

Sanford, N., Veckenstedt, R., & Moritz, S. (2014). Impaired integration of disambiguating evidence in delusional schizophrenia patients. *Psychological Medicine*, *44*, 2729–2738. Retrieved from http://journals.cambridge.org/article_S0033291714000397

Schwartz, M. (1982). Repetition and Rated Truth Value of Statements. *The American Journal of Psychology*, *95*(3), 393–407. https://doi.org/10.2307/1422132

Schwarz, N., Sanna, L. 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*(6), 127–161. https://doi.org/10.1016/S0065-2601(06)39003-X

Shane, S. (2017). From Headline to Photograph, a Fake News Masterpiece. Retrieved March 2, 2017, from https://www.nytimes.com/2017/01/18/us/fake-news-hillary-clinton-cameron-harris.html

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

Singal, J. (2017). This is a great psychological framework for understanding how fake news spreads. Retrieved from http://nymag.com/scienceofus/2017/01/a-great-psychological-framework-for-understanding-fake-news.html

Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How Warnings about False Claims Become Recommendations. *Journal of Consumer Research*, *31*(4), 713–724. https://doi.org/10.1086/426605

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Swire, B., Ecker, U. K. H., & Lewandowsky, S. (2017). The Role of Familiarity in Correcting Inaccurate Information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000422

Tauber, S., Dunlosky, J., & Rawson, K. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research*. Retrieved from http://link.springer.com/article/10.3758/s13428-012-0307-9

Thompson, V. A., Ackerman, R., Sidi, Y., Ball, L. J., Pennycook, G., & Prowse Turner, J. A. (2013). The role of answer fluency and perceptual fluency in the monitoring and control of reasoning: Reply to Alter, Oppenheimer, and Epley (2013). *Cognition*, *128*(2), 256–258. https://doi.org/10.1016/j.cognition.2013.03.003

Thompson, V. A. V. A., Prowse Turner, J. A., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Unkelbach, C. (2006). The Learned Interpretation of Cognitive Fluency. *Psychological Science, 17*(4), 339–345. https://doi.org/10.1111/j.1467-9280.2006.01708.x

Unkelbach, C. (2007). Reversing the Truth Effect : Learning the Interpretation of Processing Fluency in Judgments of Truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Unkelbach, C., & Rom, S. (2017). A referential theory of experienced truth. *Cognition*, *160*, 110–126. https://doi.org/10.1017/CBO9781107415324.004

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

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect - the Panas Scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Whittlesea, B. W. a. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253. https://doi.org/10.1037/0278-7393.19.6.1235

Yonelinas, A. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0749596X02928640

Yonelinas, A., & Jacoby, L. (2012). The process-dissociation approach two decades later: Convergence, boundary conditions, and new directions. *Memory & Cognition*. Retrieved from http://link.springer.com/article/10.3758/s13421-012-0205-5

Zajonc, R. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social*

Psychology. Retrieved from http://psycnet.apa.org/journals/psp/9/2p2/1/

Zajonc, R. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science*. Retrieved from http://journals.sagepub.com/doi/abs/10.1111/1467-8721.00154

Supplementary Information

for

Implausibility and illusory truth: Prior exposure increases perceived accuracy of fake news but has no effect for entirely implausible statements

Gordon Pennycook, Tyrone D. Cannon, & David G. Rand

Contents

1. Study 2 – Further analysis of accuracy judgments... 2

2. Study 3 – Further analysis of accuracy judgments, Session 1 4

3. Study 3 – Further analysis of accuracy judgments, Session 2 6

4. Study 3 – Social media sharing 12

5. Materials (Study 1) 13

1. Study 2 – Further analysis of accuracy judgments

Our preregistration called for a target sample size of 500. However, we decided to complete a full replication of the results with another 500 participants. Since both experiments yielded very similar results (see Table S1), they were combined in the main text.

Perceived accuracy was entered into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) x 2 (Sample: Original sample, replication) mixed ANOVA. There was no significant main effect of sample, $F < 1$, but there was a marginally significant interaction between sample and exposure, $F(1, 945) = 3.65$, $p = .056$, $\eta^2 = .004$, such that the repetition effect was somewhat larger in the replication (although the exposure effect is significant in every case, see Table S1). There were no further significant interactions between sample and other factors in the experiment, all *F*'s < 1.

Table S1 – Study 2, Original and Replication Samples. Means, standard deviations, and significance tests (comparing familiarized and novel items) for fake and real news accuracy judgments as a function of warning manipulation.

	Type	Warning	Familiarized	Novel	*t* (df)	*p*
Original Sample	Fake News	No Warning	1.81 (0.6)	1.71 (0.5)	3.14 (247)	.002
		Warning	1.68 (0.5)	1.59 (0.5)	3.17 (236)	.002
	Real News	No Warning	2.74 (0.5)	2.64 (0.5)	4.14 (247)	< .001
		Warning	2.69 (0.5)	2.63 (0.5)	1.98 (236)	.049
Replication	Fake News	No Warning	1.84 (0.6)	1.67 (0.5)	5.07 (239)	< .001
		Warning	1.73 (0.6)	1.62 (0.5)	3.74 (223)	< .001
	Real News	No Warning	2.74 (0.6)	2.59 (0.6)	4.77 (239)	< .001
		Warning	2.72 (0.6)	2.63 (0.6)	3.06 (223)	.003

We also used a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA with perceived accuracy for only those items that participants indicated being unfamiliar with (see Table S2 for means). This revealed a significant effect of exposure, $F(1, 693) = 27.67$, $p < .001$, $\eta^2 = .04$, wherein familiarized items were rated as more accurate than novel items. There was also a main effect of type, $F(1, 693) = 1091.76$, $p < .001$, $\eta^2 = .61$, such that real headlines were rated as more accurate than fake headlines. No other main effects or interactions were significant, all *F*'s < 2.3, *p*'s > .130.

Table S2 – Study 2, Unfamiliar Items. Means, standard deviations, and significance tests (comparing familiarized and novel items) for fake and real news accuracy judgments as a function of warning manipulation.

Type	Warning	Familiarized	Novel	*t* (df)	*p*
Fake News	No Warning	1.67 (0.6)	1.62 (0.5)	1.73 (443)	.084
	Warning	1.62 (0.6)	1.55 (0.5)	2.78 (343)	.006
Real News	No Warning	2.44 (0.6)	2.31 (0.6)	4.12 (410)	< .001
	Warning	2.42 (0.7)	2.34 (0.6)	2.51 (383)	.012

Finally, we report the full data for the familiarity manipulation check. For this, we entered reported familiarity into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed design ANOVA (see Table S3 for means). As reported in the main text, there was a main effect of exposure such that familiarized headlines were rated as more familiar than novel headlines, *F*(1, 947) = 578.76, *p* < .001, *η2* = .38. The analysis also revealed a main effect of type such that real headlines were rated as more familiar than fake headlines, *F*(1, 947) = 594.10, *p* < .001, *η2* = .39. There was no main effect of warning, *F*(1, 947) = 1.90, *p* = .169, *η2* = .002, or interaction between warning and type, *F*(1, 947) = 1.01, *p* = .316, *η2* = .001. There was a marginally significant interaction between exposure and warning, *F*(1, 947) = 3.64, *p* = .057, *η2* = .004, and a reliable interaction between type and exposure, *F*(1, 947) = 12.61, *p* < .001, *η2* = .01. Finally, there was a three-way interaction between type, exposure, and warning, *F*(1, 947) = 37.40, *p* < .001, *η2* = .04. To decomposing this three-way interaction, we computed a "familiarity effect" variable by subtracting the proportion of familiar items that were familiarized from those that were novel (separately for fake and real). We then entered these difference scores into a 2 (Type: fake, real) x 2 (Warning: warning, no warning) mixed ANOVA. There was a significant interaction between type and warning, *F*(1, 947) = 37.40, *p* < .001, *η2* = .04, indicating a larger influence of prior exposure on familiarity judgments in the warning condition. As is evident from Table S3, this was driven primarily by fake news items.

Table S3 – Study 2, Familiarity. Means, standard deviations, and significance tests (comparing familiarized and novel items) for fake and real news familiarity (proportion saying "no" to the familiarity question) as a function of warning manipulation.

Type	Warning	Familiarized	Novel	*t* (df)	*p*
Fake News	No Warning	.34 (0.4)	.09 (.1)	14.96 (487)	< .001
	Warning	.41 (0.4)	.05 (0.1)	17.3 (460)	< .001
Real News	No Warning	.51 (0.4)	.24 (0.2)	16.6 (487)	< .001
	Warning	.53 (0.3)	.27 (0.3)	14.1 (460)	< .001

2. Study 3 – Further analysis of accuracy judgments, Session 1

In the main text, we report the mixed ANOVA analysis using only cases where participants indicated being previously unfamiliar with the news headlines (i.e., prior to the study). The following is the same 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed ANOVA but using the full set of data. Most importantly, headlines presented in the familiarization stage were rated as more accurate than novel headlines (see Table S4 for all means), $F(1, 938) = 18.98$, $p < .001$, $\eta^2 = .02$. There was also a significant main effect of type, such that real stories were rated as much more accurate than fake stories, $F(1, 938) = 2065.63$, $p < .001$, $\eta^2 = .69$. Unlike for the restricted analysis, there was a marginally significant interaction between exposure and type of news headline, $F(1, 938) = 4.16$, $p = .042$, $\eta^2 = .004$. Nonetheless, the overall exposure effect was robust for fake news, $t(938) = 4.71$, $p < .001$, $d = .15$. The exposure effect was not significant for real news, $t(938) = 1.58$, $p = .115$, $d = .05$, perhaps due to high familiarity of the news headlines prior to the experiment (participants were unfamiliar with 81.2% of the fake news headlines, but only 49.2% of the real news headlines).

Table S4 – Study 3, Session 1. Means, standard deviations, and significance tests (comparing familiarized and novel items) for fake and real news accuracy judgments as a function of warning manipulation.

	Type	Warning	Familiarized	Novel	*t* (df)	*p*
All items	Fake News	No Warning	1.88 (0.6)	1.74 (0.6)	5.31 (475)	< .001
		Warning	1.78 (0.6)	1.75 (0.5)	1.30 (463)	.193
	Real News	No Warning	2.72 (0.6)	2.67 (0.7)	2.01 (475)	.045
		Warning	2.71 (0.6)	2.70 (0.6)	0.28 (463)	.777
Previously unfamiliar items	Fake News	No Warning	1.71 (0.6)	1.58 (0.5)	5.41 (454)	< .001
		Warning	1.63 (0.6)	1.55 (0.5)	3.07 (447)	.002
	Real News	No Warning	2.33 (0.7)	2.26 (0.7)	1.68 (373)	.094
		Warning	2.34 (0.7)	2.30 (0.7)	1.00 (357)	.316

There was no main effect of the warning manipulation, $F < 1$, and the interaction between type of news story and warning condition was not reliable, $F(1, 938) = 2.62$, $p = .106$, $\eta^2 = .003$. Nonetheless, familiarized fake news items (i.e., the fake news stories that were warned about in the familiarization stage) were rated as less accurate ($M = 1.88$, $SD = .61$) than the same stories in the control (no warning) condition ($M = 1.78$, $SD = .58$), $t(938) = 2.60$, $p = .009$, $d = .17$. There were no significant differences between the warning and no warning conditions for real news or novel fake news, all *t*'s < 1.

Unlike in the analysis with only previously unfamiliar items, there was a significant interaction between exposure and warning for the full set of data, $F(1, 938) = .81$, $p = .005$, $\eta^2 = .008$. However, there was no three-way interaction between warning, exposure, and news type, $F(1, 938) = 1.22$, $p = .269$, $\eta^2 = .001$. The overall repetition effect (for both fake and real news) was evident in the no warning condition, $t(475) = 5.04$, $p < .001$, $d = .22$, but not the warning condition, *t*(463) = 1.11, *p* = .268, *d* = .05 (see Table S4 for separate analyses of real and fake news; see also footnote 7 in main text).

3. Study 3 – Further analysis of accuracy judgments, Session 2

We also investigated whether the repetition effect would persist after a week. Participants were invited (via email) to complete another survey after a week passed. The follow-up consisted of the 16 items from the first session plus 8 new items (half real, half fake; counterbalanced between-subject). We had full data for 566 participants (60.2% of the original sample). To verify that the participants who returned did not differ from those who did not return on the crucial test, we re-ran the full mixed ANOVA with accuracy judgments for unfamiliar items. There were no significant differences between the two groups (i.e., no main effect of return/no return and no interactions between return/no return and other factors in the model), all *F*'s < 2.3, *p*'s > .130.

We analyzed the data in two different ways. First, we classified both sets of items presented in the first session as "familiarized" and contrasted them with the novel items not presented in the first session. Perceived accuracy was entered into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed ANOVA. Participants successfully rated fake news as less accurate than real news, *F*(1, 564) = 1365.66, *MSE* = .40, *p* < .001, η^2 = .71. Crucially, there was a robust main effect of exposure such that familiarized items were rated as more accurate than novel items (see Table S5), *F*(1, 564) = 56.88, *MSE* = .12, *p* < .001, η^2 = .09. There was no main effect of warning, *F* < 1, or significant interactions, all *F*'s < 1. Parallel results were found when the analysis was isolated to previously unfamiliar items (see Table S5): Fake news was rated as less accurate than real news, *F*(1, 453) = 760.17, *MSE* = .42, *p* < .001, η^2 = .63, familiarized items were rated as more accurate than novel items, *F*(1, 453) = 12.91, *MSE* = .21, *p* < .001, η^2 = .03, and there were no further effects, all *F*'s < 1.3, *p*'s > .260. However, these results should be interpreted with caution since items introduced in the second session were rated as more familiar than those introduced in the first session, *t*(565) = 4.19, *p* < .001 (which is not surprising, given that participants had an extra week in which to become familiar with headlines). Thus, a greater number of familiar items are removed from the novel items, which artificially decreases mean perceived accuracy (i.e., since familiar items are considered more accurate). Nonetheless, there was a significant effect of exposure in both cases.

Second, we investigated repetition effects from a dose-response perspective (this is the analysis we focus on in the main text). To do so, we took into account the fact that among the familiarized stories from the first session, those shown in both the familiarization stage and the assessment stage had been seen twice by participants, whereas those shown only in the assessment stage had been seen once. Thus, we have three levels of experimental exposure: there are stories in our dataset which we showed to participants either zero, one or two times. To take this into account in an analysis, we predicted average accuracy ratings for each subject for each of these three different groups of headlines, as a function of the number of times each group of headlines had been exposed. Specifically, we used linear regression with three average accuracy rating observations per participant (for headlines seen 0 times, 1 time, and 2 times), taking number of times seen as the independent variable and clustering standard errors on participant to account for the multiple observations per participant. Full regression tables for analyses including all items are presented in Tables S6 and S7, and for analyses using only previously unfamiliar items in Tables S8 and S9. Across all specifications, we find clear support for a repetition effect for both fake and real news that is increasing in the number of exposures participants experienced.

Table S5 – Study 3, Session 2. Means, standard deviations, and significance tests (comparing familiarized and novel items) for fake and real news accuracy judgments as a function of warning manipulation.

	Type	Warning	Familiarized Twice	Familiarized Once	Novel	*F* (df)	*p*
All items	Fake News	No Warning	1.91 (0.6)	1.83 (0.6)	1.73 (0.5)	13.12 (2, 536)	< .001
		Warning	1.81 (0.6)	1.80 (0.6)	1.72 (0.5)	5.21 (2, 592)	.006
	Real News	No Warning	2.82 (0.6)	2.83 (0.6)	2.72 (0.6)	6.02 (2, 536)	.003
		Warning	2.83 (0.6)	2.79 (0.5)	2.70 (0.6)	8.73 (2, 592)	< .001
Previously unfamiliar items	Fake News	No Warning	1.73 (0.6)	1.70 (0.6)	1.61 (0.5)	5.76 (2, 498)	.003
		Warning	1.67 (0.6)	1.63 (0.6)	1.60 (0.5)	2.22 (2, 552)	.110
	Real News	No Warning	2.55 (0.7)	2.50 (0.7)	2.46 (0.7)	1.30 (2, 366)	.273
		Warning	2.57 (0.7)	2.48 (0.7)	2.39 (0.7)	4.71 (2, 406)	.010

Table S6 – Regressions for Study 3, Session 2 all fake news items. Ordinary least squares regression with robust standard errors clustered on participant, taking average accuracy judgments as the dependent variable.

	(1) All stories	(2) Fake	(3) Fake No Warning	(4) Fake Warning	(5) Fake	(6) Fake Concordant	(7) Fake Non-concordant	(8) Fake
# of Exposures (0-2)	0.058*** (0.0090)	0.063*** (0.012)	0.082*** (0.017)	0.045** (0.017)	0.082*** (0.017)	0.081*** (0.017)	0.045** (0.016)	0.045** (0.016)
Condition (0=No Warning, 1=Warning)					-0.0018 (0.045)			
Condition X Exposures					-0.037 (0.024)			
Politically concordant (0=N, 1=Y)								0.15*** (0.034)
Concordant X Exposures								0.036 (0.023)
Constant	2.22*** (0.017)	1.71*** (0.022)	1.72*** (0.033)	1.71*** (0.030)	1.72*** (0.033)	1.79*** (0.030)	1.64*** (0.026)	1.64*** (0.026)
Observations	1,581	1,581	750	831	1,581	1,578	1,578	3,156
Clusters (# of participants)	527	527	250	277	527	526	526	526
R-squared	0.012	0.008	0.015	0.004	0.010	0.008	0.003	0.022

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S7 – Regressions for Study 3, Session 2 all real news items. Ordinary least squares regression with robust standard errors clustered on participant, taking average accuracy judgments as the dependent variable.

	(1) Real	(2) Real No Warning	(3) Real Warning	(4) Real	(5) Real Concordant	(6) Real Non-concordant	(7) Real
# of Exposures (0-2)	0.053*** (0.013)	0.042* (0.019)	0.063*** (0.016)	0.042* (0.019)	0.046** (0.016)	0.060** (0.018)	0.060** (0.018)
Condition (0=No Warning, 1=Warning)				-0.019 (0.048)			
Condition X Exposures				0.020 (0.025)			
Politically concordant (0=N, 1=Y)							0.37*** (0.035)
Concordant X Exposures							-0.014 (0.024)
Constant	2.73*** (0.024)	2.74*** (0.037)	2.72*** (0.031)	2.74*** (0.037)	2.91*** (0.030)	2.54*** (0.030)	2.54*** (0.030)
Observations	1,581	750	831	1,581	1,578	1,578	3,156
Clusters (# of participants)	527	250	277	527	526	526	526
R-squared	0.006	0.003	0.008	0.006	0.003	0.005	0.061
Robust standard errors in parentheses *** p<0.001, ** p<0.01, * p<0.05							

Table S8 – Regressions for Study 3, Session 2 unfamiliar fake news items. Ordinary least squares regression with robust standard errors clustered on participant, taking average accuracy judgments as the dependent variable.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All stories	Fake	Fake No Warning	Fake Warning	Fake	Fake Concordant	Fake Non-concordant	Fake
# of Exposures (0-2)	0.043***	0.048***	0.061**	0.036	0.061**	0.060**	0.041*	0.041*
	(0.012)	(0.013)	(0.019)	(0.018)	(0.019)	(0.019)	(0.018)	(0.018)
Condition (0=No Warning, 1=Warning)					-0.023			
					(0.045)			
Condition X Exposures					-0.025			
					(0.026)			
Concordant (0=N, 1=Y)								0.11**
								(0.035)
Concordant X Exposures								0.019
								(0.026)
Constant	1.90***	1.61***	1.62***	1.60***	1.62***	1.66***	1.55***	1.55***
	(0.021)	(0.022)	(0.033)	(0.031)	(0.033)	(0.031)	(0.026)	(0.026)
Observations	1,581	1,581	750	831	1,581	1,488	1,548	3,036
Clusters (# of participants)	527	527	250	277	527	524	526	526
R-squared	0.005	0.005	0.008	0.003	0.007	0.004	0.002	0.011

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table S9 – Regressions for Study 3, Session 2 unfamiliar real news items. Ordinary least squares regression with robust standard errors clustered on participant, taking average accuracy judgments as the dependent variable.

	(1) Real	(2) Real No Warning	(3) Real Warning	(4) Real	(5) Real Concordant	(6) Real Non-concordant	(7) Real
# of Exposures (0-2)	0.055**	0.024	0.083**	0.024	0.052*	0.030	0.030
	(0.020)	(0.030)	(0.026)	(0.030)	(0.026)	(0.025)	(0.025)
Condition (0=No Warning, 1=Warning)				-0.054			
				(0.064)			
Condition X Exposures				0.059			
				(0.040)			
Concordant (0=N, 1=Y)							0.26***
							(0.050)
Concordant X Exposures							0.022
							(0.036)
Constant	2.44***	2.47***	2.42***	2.47***	2.60***	2.34***	2.34***
	(0.032)	(0.048)	(0.042)	(0.048)	(0.042)	(0.038)	(0.038)
Observations	1,390	655	735	1,390	1,106	1,186	2,292
Clusters (# of participants)	521	245	276	521	497	501	520
R-squared	0.004	0.001	0.009	0.005	0.003	0.001	0.030

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

4. Study 3 – Social media sharing

In addition to rating the accuracy of each headlines, in Study 3 participants were also asked to indicate whether they would be willing to share the headline on social media. In our preregistration, we indicated that willingness to share on social media will be scored 0 if "no" is selected and 1 if "maybe" or "yes" is selected. We also noted that cases where people indicate that they would never share something political online or who don't use social media will be removed.

To investigate whether the exposure induction had an effect on social media sharing, we entered the willingness to share on social media measure into a 2 (Type: fake, real) x 2 (Exposure: familiarized, novel) x 2 (Warning: warning, no warning) mixed ANOVA. Only participants who contributed data to each cell of this design were included (*N* = 546). This analysis revealed that fake news was less likely to be shared than real news, *F*(1, 544) = 201.51, *MSE* = .07, $p < .001$, $\eta^2 = .27$. Paradoxically, participants were *more* willing to share novel news headlines relative to familiarized ones, *F*(1, 544) = 9.37, *MSE* = .03, $p = .002$, $\eta^2 = .02$. However, as is evident from Table S10, none of the direct comparisons between familiarized and novel headlines were significant. Finally, the warning had only a marginally negative overall effect on social media sharing, *F*(1, 544) = 3.43, *MSE* = .22, $p = .065$, $\eta^2 = .006$. No interactions were significant, all *F*'s < 1.

Table S10 – Study 3, Social Media Sharing. Means, standard deviations, and significance tests (comparing familiar and unfamiliar items) for fake and real news as a function of source manipulation. These data include every possible participant and thus the means differ slightly than what was used in the full ANOVA reported above.

	Type	Warning	Familiarized	Novel	*t* (df)	*p*
Same session	Fake News	No Warning	.24 (.3)	.25 (.3)	0.66 (275)	.509
		Warning	.20 (.3)	.23 (.3)	1.81 (269)	.072
	Real News	No Warning	.40 (.3)	.44 (.3)	1.85 (275)	.066
		Warning	.36 (.3)	.39 (.3)	1.60 (269)	.110
One week follow-up	Fake News	No Warning	.18 (.2)	.18 (.3)	0.23 (145)	.823
		Warning	.19 (.3)	.20 (.3)	0.74 (174)	.461
	Real News	No Warning	.37 (.3)	.35 (.3)	1.38 (145)	.171
		Warning	.36 (.3)	.35 (.3)	1.09 (174)	.278

We completed the parallel analysis using the data from the one week follow-up (Table S10), again excluding participants who indicating not being willing to ever share political news online (and those who do not use social media) in the first session. Participants indicated being less willing to share fake than real news, *F*(1, 319) = 151.25, *MSE* = .06, $p < .001$, $\eta^2 = .32$. No other effects were significant, *F*'s < 2.3, *p*'s > .134. Thus, in contrast to the first session, participants were no more likely to share novel than familiar stories (and, in fact, the pattern of results was in the opposite direction; see Table S10), *F*(1, 319) = 1.20, *MSE* = .02, $p = .274$, $\eta^2 = .001$.

5. Materials (Study 1)

Table S11 – Study 1 Items.

Known	False (Extremely Implausible)	Smoking cigarettes is good for your lungs.
		The earth is a perfect square.
		Across the United States, only a total of 452 people voted in the last election.
		A single elephant weighs less than a single ant.
	True	More people live in the United States than in Malta.
		Cows are larger than sheep.
		Coffee is a more popular drink in America than goat milk.
		There are more than fifty stars in the universe.
Unknown	False	George was the name of the goldfish in the story of Pinocchio.
		Johnson was the last name of the man who killed Jesse James.
		Charles II was the first ruler of the Holy Roman Empire.
		Canopus is the name of the brightest star in the sky, excluding the sun.
		Tirpitz was the name of Germany's largest battleship that was sunk in World War II.
		John Kenneth Galbraith is the name of a well-known lawyer.
		Huxley is the name of the scientist who discovered radium.
		The Cotton Bowl takes place in Auston, Texas.
		The drachma is the monetary unit for Macedonia.
		Angel Falls is located in Brazil.
	True	The thigh bone is the largest bone in the human body.
		Bolivia borders the Pacific Ocean.
		The largest dam in the world is in Pakistan.
		Mexico is the world's largest producer of silver.
		More presidents of the United States were born in Virginia than any other state.
		Helsinki is the capital of Finland.
		Marconi is name of the inventor of the wireless radio.
		Billy the Kid's last name was Bonney.
		Tiber is the name of the river that runs through Rome.
		Canberra is the capital of Australia.