Cognition 194 (2020) 104054

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

An initial accuracy focus prevents illusory truth

Nadia M. Brashier[a,*], Emmaline Drew Eliseev[b], Elizabeth J. Marsh[b]

[a] *Department of Psychology, Harvard University, United States*
[b] *Department of Psychology & Neuroscience, Duke University, United States*

ARTICLE INFO

Keywords:
Fluency
Knowledge
Fact checking
Illusory truth

ABSTRACT

News stories, advertising campaigns, and political propaganda often repeat misleading claims, increasing their persuasive power. Repeated statements feel easier to process, and thus truer, than new ones. Surprisingly, this *illusory truth effect* occurs even when claims contradict young adults' stored knowledge (e.g., repeating *The fastest land animal is the leopard* makes it more believable). In four experiments, we tackled this problem by prompting people to behave like "fact checkers." Focusing on accuracy at exposure (giving initial truth ratings) wiped out the illusion later, but only when participants held relevant knowledge. This selective benefit persisted over a delay. Our findings inform theories of how people evaluate truth and suggest practical strategies for coping in a "post-truth world."

1. Introduction

Every day, we encounter false claims that range from banal (e.g., *Lack of sleep causes jet lag*) to dangerous (e.g., *Undocumented immigrants do not pay taxes*). Advertisements, fake news sites, and political speeches repeat these and other myths, lending them a veneer of credibility: Repeated claims seem truer than new ones (Hasher, Goldstein, & Toppino, 1977). This *illusory truth effect* occurs for product claims (e.g., *Crest toothpaste removes caffeine stains from teeth*; Johar & Roggeveen, 2007), fake headlines (e.g., *Donald Trump sent his own plane to transport 200 stranded Marines*; Pennycook, Cannon, & Rand, 2018), and sociopolitical opinions (e.g., *Providing low rent housing to those on welfare only encourages these people not to work*; Arkes, Hackett, & Boehm, 1989).

Regardless of the topic at hand, repeated statements feel easy to process, or *fluent*, which people interpret as evidence of truth (Wang, Brashier, Wing, Marsh, & Cabeza, 2016). Fluency provides a strong metacognitive signal (Alter & Oppenheimer, 2009), making the illusion very difficult to wipe out. In fact, the effect is larger than initial estimates suggested (d = 0.49, Dechêne, Stahl, Hansen, & Wänke, 2010) – most researchers caution participants that they will encounter true and false claims, an instruction that cuts the illusion in half (Jalbert, Newman, & Schwarz, in preparation). Moreover, illusory truth persists in the face of sound advice (from a person labeled as 100% accurate, Unkelbach & Greifeneder, 2018), after long delays (up to months, Brown & Nix, 1996), with warnings (Nadarevic & Aßfalg, 2016), and despite explicit indications that claims come from untrustworthy sources (Henkel & Mattson, 2011). The illusion is also immune to individual differences in fluid intelligence and cognitive style (De Keersmaecker et al., in press).

The picture does not improve when we consider knowledge of specific facts, rather than general intellect. Intuitively, repeatedly contradicting a well-known fact (e.g., *The fastest land animal is the leopard*) should not make it believable. But surprisingly, it does – illusory truth occurs even when people "know better" (e.g., that the cheetah, not the leopard, is fastest; Fazio, Brashier, Payne, & Marsh, 2015). Repetition may even inflate belief in highly implausible statements, like *The Earth is a perfect square* (Fazio, Rand, & Pennycook, in press). In the current climate of misinformation, where false news travels further and faster than the truth (Vosoughi, Roy, & Aral, 2018), not only the uninformed are at risk.

Crucially, the problem is not intractable – repeating *The fastest land animal is the leopard* does not sway older adults (Brashier, Umanath, Cabeza, & Marsh, 2017). Young adults may simply need a nudge to prioritize accuracy. As examples, they choose to share news they recognize as false (Pennycook et al., 2018; Pew Research Center, 2016), tell frequent lies (Serota, Levine, & Boster, 2010), and fall for fake headlines because of "lazy thinking" (Pennycook & Rand, 2019). But asking people to behave like "fact checkers" creates an accuracy focus. Fact checking often refers to an external process (consulting another person or Google; e.g., Risko, Ferguson, & McLean, 2016), but it can also occur internally (searching memory). For example, after striking through and correcting errors in a text (e.g., *Wearing a seatbelt decreases the likelihood of surviving a car accident*), participants judge these assertions to be less truthful later (Rapp, Hinze, Kohlhepp, & Ryskin,

* Corresponding author at: Department of Psychology, Harvard University, 33 Kirkland St, Cambridge, MA 02138, United States.
E-mail address: nbrashier@fas.harvard.edu (N.M. Brashier).

https://doi.org/10.1016/j.cognition.2019.104054
Received 4 February 2019; Received in revised form 15 August 2019; Accepted 16 August 2019

2014). In the current research, we investigated whether internal fact checking, implemented as initial truth judgments, protects people from illusory truth when they "know better." Previous studies asked participants to evaluate truth at exposure (e.g., Arkes et al., 1989; Boehm, 1994; Nadarevic & Erdfelder, 2014), but used ambiguous materials (eliminating the role of previous knowledge).

In four experiments, we tested a promising strategy: asking young adults to fact check claims at exposure, thereby activating their knowledge. Participants initially rated statements for either interest or truth, then judged the truth of these statements and new items. Knowledge was defined by norms (Experiments 1 and 2), as well as individuals' performance on a post-experimental knowledge check (Experiments 3 and 4). All data and materials are available online (osf.io/b4szp/).

2. Experiment 1

2.1. Method

2.1.1. Participants

One hundred and three Amazon Mechanical Turk workers (52 female; *M* age = 36.82 years) participated for compensation. We also excluded one participant who reported looking up answers.

2.1.2. Design

This experiment had a 2 (initial rating: interest, truth) × 2 (repetition: repeated, new) mixed design. Initial rating type was manipulated between subjects, while repetition was manipulated within subjects.

2.1.3. Materials

We selected 60 facts from Tauber, Dunlosky, Rawson, Rhodes, and Sitzman (2013) general knowledge norms that were likely to be known (on average, recalled correctly by 60% of norming participants), then took an additional 60 items of similar difficulty from Wang and colleagues (2016). We were most interested in how people evaluate false claims in their environment, so we converted facts (e.g., *The fastest land animal is the cheetah*) into false statements by referring to plausible, but incorrect, alternatives (e.g., *The fastest land animal is the leopard*). To prevent response bias, we included an equal number of true fillers. We divided the statements into four sets of 30 items. Two sets appeared as falsehoods (i.e., critical items) and the other two appeared as truths (i.e., fillers) for all participants. One set of falsehoods repeated across exposure and truth rating phases, whereas the other appeared for the first time at test. Repetition was counterbalanced across participants for falsehoods.

2.1.4. Procedure

After giving informed consent, participants completed the *exposure phase*. Depending on condition, participants rated 60 statements for either interest, on a scale from 1 (*very uninteresting*) to 6 (*very interesting*), or truthfulness, using a scale from 1 (*definitely false*) to 6 (*definitely true*). They read that some statements were true and others false.

Immediately after exposure, participants completed the *truth rating phase*. In addition to the warning about true and false statements, participants read that some statements appeared earlier in the experiment, while others were new. They also received instructions not to worry about matching their previous ratings. Participants rated 120 statements for truthfulness, using a scale from 1 (*definitely false*) to 6 (*definitely true*).

2.2. Results

The alpha level was set at .05 for all statistical tests. As discussed above, analyses focused on responses to falsehoods (i.e., critical items). For each experiment, planned comparisons tested whether illusory truth varied with initial rating type.

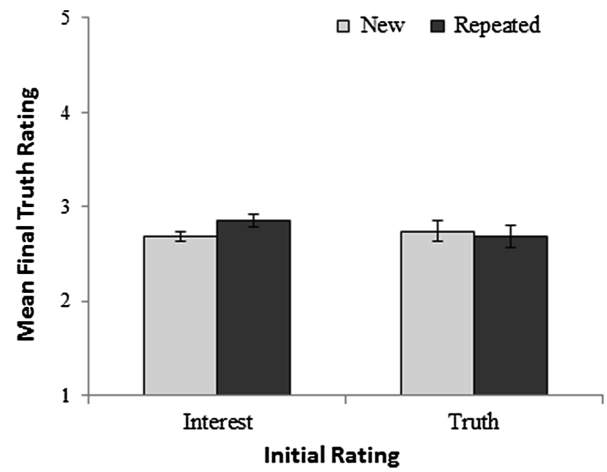

Fig. 1. Mean truth ratings for falsehoods as a function of initial rating type and repetition in Experiment 1. Error bars reflect standard error of the mean.

We conducted a 2 (initial rating: interest, truth) × 2 (repetition: repeated, new) mixed ANOVA on participants' final truth ratings for falsehoods. The relevant data appear in Fig. 1. Main effects of initial rating ($F < 1$) and repetition, $F(1, 101) = 1.61$, $p = .207$, were not significant. Critically, we found an interaction between initial rating and repetition, $F(1, 101) = 5.70$, $p = .019$, $\eta_p^2 = 0.05$. For participants in the standard condition (initial interest ratings), repeated falsehoods ($M = 2.85$) received higher final truth ratings than new ones ($M = 2.69$), $t(51) = 2.40$, $p = .020$, $d = 0.33$. For participants with an initial accuracy focus (initial truth ratings), however, repeated ($M = 2.69$) and new ($M = 2.74$) falsehoods received similar final ratings, $t(50) = 0.87$, $p = .390$.

2.3. Discussion

After a standard exposure task (initial interest ratings), participants neglected their knowledge when they could make inferences based on fluency. In contrast, an accuracy focus (initial truth ratings) wiped out the small illusion later, presumably by activating knowledge. However, this benefit could simply reflect more time spent thinking about claims (*M* reaction times: initial truth rating = 5.63 s; initial interest rating = 4.09 s) or general skepticism. To rule out these possibilities, Experiment 2 included difficult items that participants are unlikely to know. Fact checking should only reduce illusory truth for well-known claims.

3. Experiment 2

3.1. Method

3.1.1. Participants

Ninety-nine Duke University undergraduates (66 female; *M* age = 21.44 years) participated for compensation.

3.1.2. Design

This experiment had a 2 (initial rating: interest, truth) × 2 (repetition: repeated, new) × 2 (knowledge: known, unknown) mixed design. Initial rating type was manipulated between subjects, while repetition and knowledge were manipulated within subjects.

3.1.3. Materials

We selected 104 facts from Tauber and colleagues (2013) general knowledge norms. Half were likely to be known (on average, recalled by 58% of norming participants) and half were likely to be unknown (recalled by only 1% of norming participants); then we added 96

Table 1
Sample statements and multiple-choice questions.

	Statement	Knowledge check
Likely Known	Deer meat is called veal	What is the name for deer meat? (*venison*)
	The largest ocean on Earth is the Atlantic	What is the largest ocean on Earth? (*Pacific*)
	The fastest land animal is the leopard	What is the fastest land animal? (*cheetah*)
Likely Unknown	The twenty-first U.S. president was Garfield	Who was the twenty-first U.S. president? (*Arthur*)
	The author of “Brothers Karamazov” is Tolstoy	Who is the author of "Brothers Karamazov“? (*Dostoyevsky*)
	Billy the Kid's last name is Garrett	What is Billy the Kid’s last name? (*Bonney*)

Notes. The correct answer to each multiple-choice question (Experiments 3 and 4) appears in parentheses.

similar items from Wang and colleagues (2016). See Table 1 for sample statements. We divided the statements into four sets of 50 items. Two sets appeared as falsehoods (i.e., critical items) and the others appeared as truths (i.e., fillers) for all participants. One set of falsehoods repeated across exposure and truth rating phases, whereas the other appeared for the first time during the truth rating phase. Repetition was counterbalanced across participants for falsehoods.

3.1.4. Procedure

The procedure was identical to that of Experiment 1, with the exceptions that participants (a) judged more claims during the exposure (100 statements) and truth rating (200 statements) phases and (b) provided binary (*true*, *false*) final truth ratings.

3.2. Results

Participants made binary ratings at test, so we analyzed the proportion of falsehoods rated “true.”

We conducted a 2 (initial rating: interest, truth) × 2 (estimated knowledge: known, unknown) × 2 (repetition: repeated, new) mixed ANOVA on the proportion of claims participants judged to be “true” in the final phase. The relevant data appear in Fig. 2. Replicating the standard illusory truth effect, repeated falsehoods ($M = 0.56$) were more likely to be judged “true” than new falsehoods ($M = 0.49$), $F(1, 97) = 30.57$, $p < .001$, $\eta_p^2 = 0.24$. Unsurprisingly, falsehoods that contradicted well-known facts ($M = 0.48$) were less likely to be judged “true” than contradictions of unknown ones ($M = 0.57$), $F(1, 97) = 25.59$, $p < .001$, $\eta_p^2 = 0.21$. Overall, participants with an initial accuracy focus (initial truth ratings) ($M = 0.49$) made fewer “true” judgments than those in the standard (initial interest ratings) condition ($M = 0.56$), $F(1, 97) = 4.02$, $p = .048$, $\eta_p^2 = 0.04$. They were also less vulnerable to fluency (repeated $M = 0.50$, new $M = 0.48$) than participants in the standard condition (repeated $M = 0.62$, new $M = 0.50$)

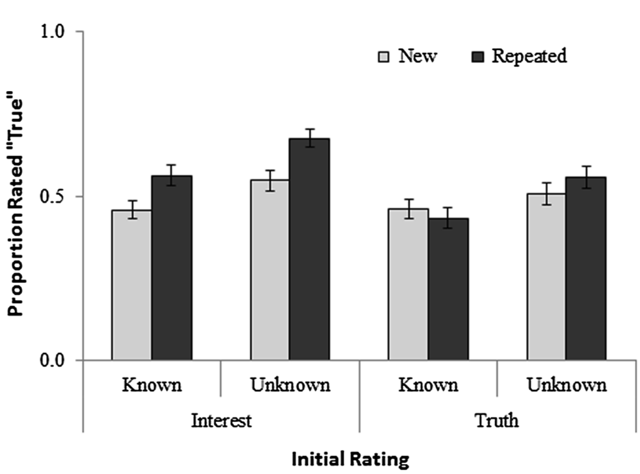

Fig. 2. Proportion of falsehoods rated “true” as a function of initial rating type, knowledge, and repetition in Experiment 2. Error bars reflect standard error of the mean.

were, $F(1, 97) = 19.87$, $p < .001$, $\eta_p^2 = 0.17$. There was no interaction between initial rating condition and knowledge, $F < 1$.

The three-way interaction among initial rating, knowledge, and repetition was not significant, $F(1, 97) = 2.29$, $p = .133$. However, the pattern of means suggested that an accuracy focus only benefited judgments of known items later. Participants in the standard condition (initial interest ratings) demonstrated illusory truth for both unknown (repeated $M = 0.67$; new $M = 0.55$, $t(48) = 5.28$, $p < .001$, $d = 0.76$) and known (repeated $M = 0.56$; new $M = 0.46$, $t(48) = 5.04$, $p < .001$, $d = 0.72$) falsehoods. Participants who made initial truth ratings exhibited illusory truth for unknown (repeated $M = 0.56$; new $M = 0.51$, $t(49) = 2.61$, $p = .012$, $d = 0.37$) but not known (repeated $M = 0.43$; new $M = 0.46$, $t(49) = 1.41$, $p = .165$) falsehoods.

3.3. Discussion

After initially rating interest, participants defaulted to fluency at test, regardless of their knowledge. An initial accuracy focus (initial truth ratings) selectively benefited known items; illusory truth only occurred for unknown items, where participants had no knowledge to activate at exposure. Initial truth ratings encourage “deeper” encoding, but this additional processing was unhelpful on its own (without knowledge). Indeed, effects of repetition *increase* with elaborative encoding: Relative to a shallow exposure task (reporting statements’ location on the screen), self-referential processing (relating statements to personal events or feelings) enhances illusory truth (Unkelbach & Rom, 2017). We note that the three-way interaction was not statistically significant, perhaps because norms only roughly estimate knowledge. Experiment 3 increased power by measuring which facts each participant knew.

4. Experiment 3

4.1. Method

4.1.1. Participants

Sixty-eight Duke University undergraduates (35 female; *M* age = 19.31 years) participated for course credit.

4.1.2. Design

This experiment had the same design as Experiment 2, except that knowledge varied within subjects.

4.1.3. Materials

We used the same statements as in Experiment 2. The final knowledge check consisted of multiple-choice questions about the falsehoods. The three answer options included the correct answer, the target misinformation presented earlier, and a *don’t know* option. For example, the question *What is the fastest land animal?* was accompanied by *cheetah*, *leopard*, and *don’t know* answer choices (see Table 1). For each participant, we categorized items as *known* or *unknown* based on recognition during the knowledge check.

4.1.4. Procedure

The procedure was identical to that of Experiment 2, with the exceptions that (a) participants assigned continuous truth ratings using a 6-point scale (as in Experiment 1) and (b) they completed a knowledge check. The final *knowledge check* included 100 multiple-choice questions with three response options: the correct answer, the alternative embedded in the falsehood seen earlier, and *don't know*. The experimenter asked participants to indicate *don't know* instead of guessing.

4.2. Results

4.2.1. Knowledge check

We first assessed knowledge check performance to ensure that our materials spanned a range of difficulty. Overall, participants answered 44% of the knowledge check questions correctly (known items). They gave wrong answers to 8% of the questions and responded to another 48% with *don't know*. Collapsing across these response types, 56% of the items were unknown. The high *don't know* rate indicates that correct answers corresponded to actual knowledge, rather than guesses. If anything, we underestimated people's knowledge. Viewing the false version of a statement may bias participants to choose the wrong answer later (Bottoms, Eslick, & Marsh, 2010; Kamas, Reder, & Ayers, 1996), working against our hypothesis.

4.2.2. Truth ratings

We conducted a 2 (initial rating: interest, truth) × 2 (demonstrated knowledge: known, unknown) × 2 (repetition: repeated, new) mixed ANOVA on participants' final truth ratings for falsehoods. The number of known and unknown items varied for each participant, depending on their knowledge check performance (minimum trials per cell = 11, *M* trials per cell = 25). The relevant data appear in Fig. 3.

Again, repeated falsehoods ($M = 3.44$) received higher truth ratings than new falsehoods ($M = 3.28$), $F(1, 66) = 5.42$, $p = .023$, $\eta_p^2 = 0.08$. As expected, falsehoods that contradicted well-known facts ($M = 2.60$) received lower (i.e., more accurate) truth ratings than contradictions of unknown ones ($M = 3.98$), $F(1, 66) = 352.81$, $p < .001$, $\eta_p^2 = 0.84$. Overall, participants with an initial accuracy focus (initial truth ratings) ($M = 3.23$) used the final truth rating scale more cautiously than those in the standard (initial interest ratings) condition ($M = 3.48$), $F(1, 66) = 6.24$, $p = .015$, $\eta_p^2 = 0.09$; they (known $M = 2.38$, unknown $M = 3.94$) also applied their knowledge more consistently than those in the standard condition (known $M = 2.80$, unknown $M = 4.01$) did, $F(1, 66) = 6.10$, $p = .016$, $\eta_p^2 = 0.09$. Finally, they were less vulnerable to fluency (repeated $M = 3.26$, new $M = 3.20$) than participants in the standard condition (repeated M = 3.60, new M = 3.35) were, $F(1, 66) = 4.03$, p = .049, $\eta_p^2 = 0.06$. There was no interaction between repetition and knowledge, $F < 1$.

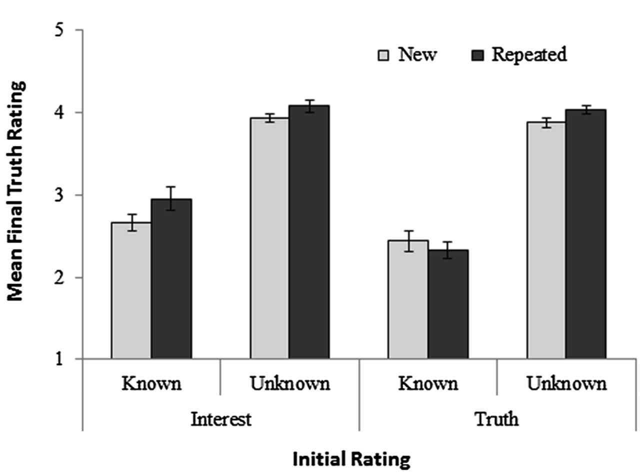

Fig. 3. Mean truth ratings for falsehoods as a function of initial rating type, knowledge, and repetition in Experiment 3. Error bars reflect standard error of the mean.

The three-way interaction among initial rating, knowledge, and repetition was significant, $F(1, 66) = 8.13$, $p = .006$, $\eta_p^2 = 0.11$. Participants in the standard condition (initial interest ratings) demonstrated illusory truth for both unknown (repeated $M = 4.07$; new $M = 3.93$, $t(34) = 2.23$, $p = .033$, $d = 0.38$) and known (repeated $M = 2.95$; new $M = 2.66$, $t(34) = 2.55$, $p = .016$, $d = 0.43$) falsehoods. Participants who made initial truth ratings exhibited illusory truth for unknown (repeated $M = 4.02$; new $M = 3.87$, $t(32) = 2.82$, $p = .008$, $d = 0.49$) but not known (repeated $M = 2.32$; new $M = 2.44$, $t(32) = 1.20$, $p = .238$) falsehoods.

4.3. Discussion

Across three experiments, focusing on accuracy at exposure eliminated any hint of an illusory truth effect for known items. But to be practical, this strategy must work over a delay. Repeated items continue to feel fluent as time passes (Dechêne et al., 2010), with illusory truth emerging even when three months elapse between exposures (Brown & Nix, 1996). In real life, we often encounter misinformation (e.g., *The U.S. has the cleanest air in the world*) well before we need to make important judgments (e.g., whether to vote for a proponent of the Green New Deal). To demonstrate long-term benefits of fact checking, Experiment 4 introduced a two-day delay between exposure and final truth judgments.

5. Experiment 4

5.1. Method

5.1.1. Participants

Eighty-nine Duke students (64 female; *M* age = 20.17 years) participated for course credit or monetary compensation. Three participants completed the exposure phase, but not the second session. We excluded another participant who performed poorly on the knowledge check (11% correct).

5.1.2. Design, materials, and procedure

This experiment is identical to Experiment 3, except that a delay of two days separated the exposure and truth rating phases.

5.2. Results

5.2.1. Knowledge check

Overall, participants answered 45% of the knowledge check questions correctly (known items). They gave wrong answers to 7% of the questions and responded to another 48% with *don't know*. Collapsing across these response types, 55% of the items were unknown.

5.2.2. Truth ratings

We conducted a 2 (initial rating: interest, truth) × 2 (demonstrated knowledge: known, unknown) × 2 (repetition: repeated, new) mixed ANOVA on participants' final truth ratings for falsehoods. Again, the number of known and unknown items varied for each participant (minimum trials per cell = 7, *M* trials per cell = 25). The relevant data appear in Fig. 4.

After a delay, repeated falsehoods ($M = 3.45$) still received higher truth ratings than new falsehoods ($M = 3.34$), $F(1, 87) = 8.70$, $p = .004$, $\eta_p^2 = 0.09$. As expected, falsehoods that contradicted well-known facts ($M = 2.67$) received lower (i.e., more accurate) truth ratings than contradictions of unknown ones ($M = 4.01$), $F(1, 87) = 361.21$, $p < .001$, $\eta_p^2 = 0.81$. There was no main effect of initial rating condition, $F(1, 87) = 1.58$, $p = 0.212$.

The three-way interaction among initial rating, knowledge, and repetition was significant, $F(1, 87) = 4.71$, $p = .033$, $\eta_p^2 = 0.05$.

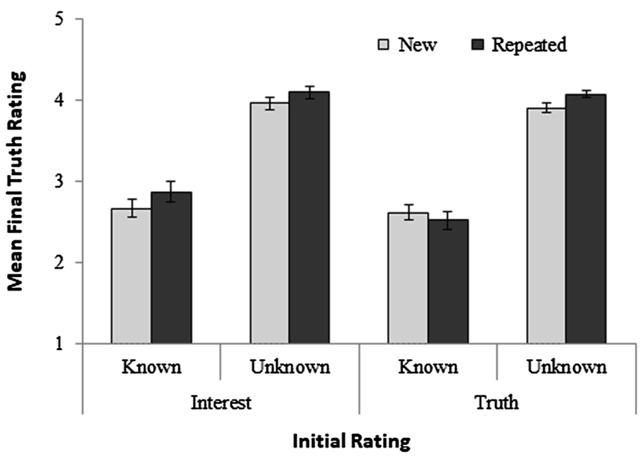

Fig. 4. Mean truth ratings for falsehoods as a function of initial rating type, knowledge, and repetition in Experiment 4. Error bars reflect standard error of the mean.

Participants in the standard condition (initial interest ratings) demonstrated illusory truth for both unknown (repeated $M = 4.10$; new $M = 3.96$, $t(45) = 2.60$, $p = .013$, $d = 0.38$) and known (repeated $M = 2.87$; new $M = 2.67$, $t(45) = 2.28$, $p = .028$, $d = 0.34$) falsehoods. Participants who made initial truth ratings exhibited illusory truth for unknown (repeated $M = 4.08$; new $M = 3.91$, $t(42) = 2.85$, $p = .007$, $d = 0.43$) but not known (repeated $M = 2.52$; new $M = 2.61$, $t(42) = 1.10$, $p = .276$) falsehoods. No other interactions were significant, $Fs < 3.47$, $ps > 0.066$.

6. General discussion

The present studies successfully prevented people from accepting fluent falsehoods as true when they "knew better." Focusing on claims' accuracy protected people from fluency later, so long as they had relevant knowledge stored in memory. Impressively, the selective benefit of fact checking persisted over two days. This finding is striking, since illusory truth stubbornly emerges over long delays, among intelligent people, for claims explicitly tagged as "false," and despite reliable advice (Brashier & Marsh, 2020).

In the face of fluency, knowledge is not always power. For example, people answer more questions containing false premises (e.g., *In the biblical story, what was Joshua swallowed by?*) when they are printed in easy-to-read (fluent) fonts; they pass over these errors, even though they know better (e.g., that the whale swallowed Jonah, not Joshua; Song & Schwarz, 2008). Similarly, our results demonstrate that young adults need a nudge to retrieve knowledge rather than using a "shortcut" to judge truth – unlike older adults who spontaneously use stored knowledge (Brashier et al., 2017). Thus, education only offers part of the solution to the misinformation crisis; we must also prompt people to carefully compare incoming claims to what they already know. Fact checking takes advantage of information already stored in memory, generalizes across domains, feels less invasive than censoring or manipulating content (see public outcry to Kramer, Guillory, & Hancock, 2014), and prevents (rather than corrects) misconceptions. This is ideal, since people often continue to believe misinformation after debunking (Chan, Jones, Jamieson, & Albarracín, 2017).

Of course, fact checking is not always efficient, given limited cognitive resources. Retrieving knowledge requires time and effort, whereas fluent judgments tend to be fast, easy, and accurate (Unkelbach, 2007). Without background knowledge, fact checking takes even longer, as people must consult external sources like Google. Carefully evaluating truth offers the most practical value ahead of high-stakes judgments (e.g., about large purchases or political candidates), where accuracy is key.

Our results bear on people's daily lives, where repetition shapes important decisions (Unkelbach, Koch, Silva, & Garcia-Marques, 2019). As examples, professors assign higher grades in subsequent offerings of the same course, and judges rate competitors more favorably in later seasons of *Dancing with the Stars* (O'Connor & Cheema, 2018). Reliance on a fluency heuristic has dire consequences in a "post-truth world" (see Lewandowsky, Ecker, & Cook, 2017), where falsehoods tend to be repeated. The most common advice is to *consider the source*, but people struggle to remember sources (Henkel & Mattson, 2011) and tagging some fake news stories as "false" boosts the perceived accuracy of untagged ones (*implied truth effect*; Pennycook, Bear, Collins, & Rand, 2019). In addition to considering the source, our findings suggest that we can simply ask ourselves *is this true?*

Supplementary material

All data and materials are available on the Open Science Framework (osf.io/b4szp).

Declaration of Competing Interest

None.

Acknowledgements

We gratefully acknowledge funding from the American Psychological Association (NMB), and American Psychological Foundation (NMB), Duke Interdisciplinary Behavioral Research Center (EDE), and James S. McDonnell Foundation (EJM). A National Science Foundation Postdoctoral Research Fellowship (1808571) supports the first author. We thank Meredith Mechanik for collecting pilot data.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*, 219–235. https://doi.org/10.1177/1088868309341564.
Arkes, H. R., Hackett, C., & Boehm, L. (1989). The generality of the relation between familiarity and judged validity. *Journal of Behavioral Decision Making, 2*, 81–94. https://doi.org/10.1002/bdm.3960020203.
Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin, 20*, 285–293. https://doi.org/10.1177/0146167294203006.
Bottoms, H. C., Eslick, A. N., & Marsh, E. J. (2010). Memory and the Moses illusion: Failures to detect contradictions with stored knowledge yield negative memorial consequences. *Memory, 18*, 670–678. https://doi.org/10.1080/09658211.2010.501558.
Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology, 71*.
Brashier, N. M., Umanath, S., Cabeza, R., & Marsh, E. J. (2017). Competing cues: Older adults rely on knowledge in the face of fluency. *Psychology and Aging, 32*, 331–337. https://doi.org/10.1037/pag0000156.
Brown, A. S., & Nix, L. A. (1996). Turning lies into truths: Referential validation of falsehoods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1088–1100. https://doi.org/10.1037/0278-7393.22.5.1088.
Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science, 28*, 1531–1546. https://doi.org/10.1177/0956797617714579.
De Keersmaecker, J., Dunning, D., Pennycook, G., Rand, D. G., Sanchez, C., Unkelbach, C., & Roets, A. (in press). Investigating the robustness of the illusory truth effect across individual differences in cognitive ability, need for cognitive closure, and cognitive style. Personality and Social Psychology Bulletin. https://doi.org/10.1177/0146167219853844.
Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*, 238–257. https://doi.org/10.1177/1088868309352251.
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*, 993–1002. https://doi.org/10.1037/xge0000098.
Fazio, L. K., Rand, D. G., & Pennycook, G. (in press). Repetition increases perceived truth equally for plausible and implausible statements. Psychonomic Bulletin & Review. https://doi.org/10.3758/s13423-019-01651-4.
Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior, 16*, 107–112. https://doi.org/10.1016/S0022-5371(77)80012-1.
Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source

credibility. *Consciousness and Cognition, 20*, 1705–1721. https://doi.org/10.1016/j.concog.2011.08.018.

Jalbert, M., Newman, E., & Schwarz, N. (submitted for publication). Only half of what I'll tell you is true: How experimental procedures lead to an underestimation of the truth effect.

Johar, G. V., & Roggeveen, A. L. (2007). Changing false beliefs from repeated advertising: The role of claim-refutation alignment. *Journal of Consumer Psychology, 17*, 118–127. https://doi.org/10.1016/S1057-7408(07)70018-9.

Kamas, E. N., Reder, L. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition, 24*, 687–699. https://doi.org/10.3758/BF03201094.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences USA, 111*, 8788–8790. https://doi.org/10.1073/pnas.1320040111.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth era". *Journal of Applied Research in Memory and Cognition, 6*, 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008.

Nadarevic, L., & Aßfalg, A. (2016). Unveiling the truth: Warnings reduce the repetition-based truth effect. *Psychological Research, 81*, 1–13. https://doi.org/10.1007/s00426-016-0777-y.

Nadarevic, L., & Erdfelder, E. (2014). Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition, 23*, 74–84. https://doi.org/10.1016/j.concog.2013.12.002.

O'Connor, K., & Cheema, A. (2018). Do evaluations rise with experience? *Psychological Science, 29*, 779–790. https://doi.org/10.1177/0956797617744517.

Pennycook, G., Bear, A., Collins, E., & Rand, D. G. (2019). The implied truth effect: Attaching warnings to a subset of fake news stories increases perceived accuracy of stories without warnings. *Management Science*. https://doi.org/10.2139/ssrn.3035384.

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

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

Pew Research Center (2016). Many Americans believe fake news is sowing confusion. Retrieved from www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion.

Rapp, D. N., Hinze, S. R., Kohlhepp, K., & Ryskin, R. A. (2014). Reducing reliance on inaccurate information. *Memory & Cognition, 42*, 11–26. https://doi.org/10.3758/s13421-013-0339-0.

Risko, E. F., Ferguson, A. M., & McLean, D. (2016). On retrieving information from external knowledge stores: Feeling-of-findability, feeling-of-knowing and Internet search. *Computers in Human Behavior, 65*, 534–543. https://doi.org/10.1016/j.chb.2016.08.046.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research, 36*, 2–25. https://doi.org/10.1111/j.1468-2958.2009.01366.x.

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

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods, 45*, 1115–1143. https://doi.org/10.3758/s13428-012-0307-9.

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., & Greifeneder, R. (2018). Experiential fluency and declarative advice jointly inform judgments of truth. *Journal of Experimental Social Psychology, 79*, 78–86. https://doi.org/10.1016/j.jesp.2018.06.010.

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition, 160*, 110–126. https://doi.org/10.1016/j.cognition.2016.12.016.

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

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

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