AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

Manuscript version of

Knowledge Does Not Protect Against Illusory Truth

Lisa K. Fazio, Nadia M. Brashier, B. Keith Payne, Elizabeth J. Marsh

Funded by:

- James S. McDonnell Foundation
- National Science Foundation

© 2015, American Psychological Association. This manuscript is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final version of record is available via its DOI: https://dx.doi.org/10.1037/xge0000098

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

CHORUS *Advancing Public Access to Research*

Knowledge Does Not Protect Against Illusory Truth

Lisa K. Fazio[1]

Nadia M. Brashier[2]

B. Keith Payne[3]

and

Elizabeth J. Marsh[2]

[1] Vanderbilt University, [2] Duke University, [3] University of North Carolina at Chapel Hill

Author Note

Lisa K. Fazio, Department of Psychology and Human Development, Vanderbilt University; Nadia M. Brashier, Department of Psychology & Neuroscience, Duke University; B. Keith Payne, Department of Psychology, University of North Carolina at Chapel Hill; Elizabeth J. Marsh, Department of Psychology & Neuroscience, Duke University.

This research was supported by a Collaborative Activity award from the James S. McDonnell Foundation (EJM). A National Science Foundation Graduate Research Fellowship supported the second author.

Correspondence concerning this article should be addressed to Lisa K. Fazio, 230 Appleton Place #552, Vanderbilt University, Nashville, TN 37203. E-mail: lisa.fazio@vanderbilt.edu

Word Count: 5,231

Abstract

In daily life, we frequently encounter false claims in the form of consumer advertisements, political propaganda, and rumors. Repetition may be one way that insidious misconceptions, such as the belief that vitamin C prevents the common cold, enter our knowledge base. Research on the *illusory truth effect* demonstrates that repeated statements are easier to process, and subsequently perceived to be more truthful, than new statements. The prevailing assumption in the literature has been that knowledge constrains this effect (i.e., repeating the statement *The Atlantic Ocean is the largest ocean on Earth* will not make you believe it). We tested this assumption using both normed estimates of knowledge and individuals' demonstrated knowledge on a post-experimental knowledge check (Experiment 1). Contrary to prior suppositions, illusory truth effects occurred even when participants knew better. Multinomial modeling demonstrated that participants sometimes rely on fluency even if knowledge is also available to them (Experiment 2). Thus, participants demonstrated *knowledge neglect*, or the failure to rely on stored knowledge, in the face of fluent processing experiences.

Keywords: illusory truth, fluency, knowledge neglect

Knowledge Does Not Protect Against Illusory Truth

We encounter many misleading claims in our daily lives, some of which have the potential to affect important decisions. For example, many people purchase "toning" athletic shoes to improve their fitness or take preventative doses of Vitamin C to avoid contracting a cold. How do such misconceptions enter our knowledge base and inform our choices? One key factor appears to be repetition: Repeated statements receive higher truth ratings than new statements, a phenomenon called the *illusory truth effect*. Since Hasher, Goldstein, and Toppino's (1977) seminal study, cognitive, social, and consumer psychologists have replicated the basic effect dozens of times.

The illusory truth effect is robust to many procedural variations. While most studies utilize obscure trivia statements (e.g., Bacon, 1979), the effect also occurs for assertions about consumer products (Hawkins & Hoch, 1992; Johar & Roggeveen, 2007) and for sociopolitical opinions (Arkes, Hackett, & Boehm, 1989). Illusory truth occurs when people are only exposed to general topics (e.g., *hen's body temperature*), then later asked to evaluate specific statements (e.g., *The temperature of a hen's body is about 104 degrees Fahrenheit*; Begg, Armour, & Kerr, 1985; see also Arkes, Boehm, & Xu, 1991). The effect emerges after delays of minutes (e.g., Begg & Armour, 1991; Schwartz, 1982), weeks (Bacon, 1979; Gigerenzer, 1984), and months (Brown & Nix, 1996). Moreover, Gigerenzer (1984) replicated the effect outside of the laboratory setting using representative samples and naturalistic stimuli.

Recent work suggests that the ease with which people comprehend statements (i.e., *processing fluency*) underlies the illusory truth effect. Repetition makes statements easier to process (i.e., fluent) relative to new statements, leading people to the (sometimes) false conclusion that they are more truthful (Unkelbach, 2007; Unkelbach & Stahl, 2009). Indeed,

illusory truth effects arise even without prior exposure – people rate statements presented in high-contrast (i.e., easy-to-read) fonts as "true" more often than those presented in low-contrast fonts (Reber & Schwartz, 1999; Unkelbach, 2007). Fluency informs a variety of judgments (e.g., liking, confidence, frequency; see Alter & Oppenheimer, 2009; Iyengar & Lepper, 2000; Schwartz & Metcalfe, 1992; Tversky & Kahneman, 1973), likely because it serves as a valid cue in our day to day lives (Unkelbach, 2007).

Given the strong, automatic tendency to rely on fluency, when do people use other cues to evaluate truthfulness? The chief constraint on illusory truth identified in the literature is source recollection. Begg, Anas, and Farinacci (1992), for example, paired statements with "trustworthy" or "untrustworthy" voices. At test, statements previously spoken by untrustworthy voices received lower truth ratings than new items (i.e., reverse illusory truth effect). Multinomial process modeling shows that people rely on fluency only if they fail to recollect whether or not a statement came from a credible source (Unkelbach & Stahl, 2009). However, people may not spontaneously retrieve source information (R. Marsh, Landau, & Hicks, 1997); Henkel and Mattson (2011) found that after a 2-3 week delay, participants exhibited an illusory truth effect for statements that they later identified (correctly or incorrectly) as coming from an unreliable source. This result fits with the well-documented finding that much of the information in the knowledge base comes to mind relatively automatically, without the experience of reliving the original learning event (Barber, Rajaram, & Marsh, 2008; Conway, Gardiner, Perfect, Anderson, & Cohen, 1997; Dewhurst, Conway, & Brandt, 2009; Herbert & Burt, 2001, 2003, 2004; Merritt, Hirshman, Zamani, Hsu, & Berrigan, 2006).

The other assumed constraint on the illusory truth effect is pre-experimental knowledge. A recent meta-analytic review captures the literature as follows: "Statements have to be

ambiguous, that is, participants have to be uncertain about their truth status because otherwise the statements' truthfulness will be judged on the basis of their knowledge" (Dechêne, Stahl, Hansen, & Wänke, 2010, p. 239). Very few studies directly test this assumption, perhaps because it is so intuitive and widespread. Unkelbach and Stahl (2009) tested their multinomial model with obscure materials (knowledge parameter probabilities ranged from .01 to .05), because the authors presupposed, but did not test, a strong negative relationship between knowledge and illusory truth. An abundance of empirical work demonstrates that fluency affects judgments of new information, but how does fluency influence the evaluation of information already stored in memory?

Several studies indirectly examined the role of knowledge by testing domain experts: That is, do people with more knowledge about a particular topic show an illusory truth effect in that domain? Unfortunately, different studies yielded different answers to this question. For example, Srull (1983) asked self-rated car experts and non-experts to rate trivia statements about cars (e.g., *The Cadillac Seville has the best repair record of any American made automobile*). No statistics were reported, but experts produced a numerically smaller illusory truth effect than non-experts. Parks and Toth (2006) found similar results when participants rated claims about known and unknown companies (e.g., ChapStick vs. Raven's). Claims embedded in meaningful contexts (*fluent*) received higher truth ratings than those in irrelevant contexts (*disfluent*); critically, the illusory truth effect was more pronounced for unknown than known brands. In contrast to these two studies, Arkes, Hackett, and Boehm (1989) demonstrated that expertise *increased* susceptibility to the illusion. They exposed participants to statements from seven domains (e.g., food, literature, entertainment), then asked them to rank order their knowledge about these topics. Illusory truth occurred for statements from high-expertise domains, but not

for statements from low-expertise domains. Similar conclusions were drawn from a study where psychology majors and non-majors rated statements about psychology (Boehm, 1994). Psychology majors exhibited a larger illusory truth effect than non-majors, corroborating Arkes and colleagues' finding that domain knowledge can hurt rather than help.

However, these studies on expertise targeted specific facts that participants would not know, even the domain experts. In other words, they tested the effect of related knowledge, rather than that of knowledge for individual statements, so it is unclear what conclusions to draw. Only one study addressed the role of knowledge for specific facts, rather than broad domains. Unkelbach (2007) conducted a study of perceptual fluency with known (e.g., *Aristotle was a Japanese philosopher*) and unknown (e.g., *The capital of Madagascar is Toamasina*) items. Some statements appeared in high-contrast font colors (i.e., fluent), while others appeared in low-contrast font colors (i.e., disfluent). Replicating previous research (Reber & Schwarz, 1999), participants rated fluent items as "true" more often than disfluent items. The interaction between fluency and knowledge was not significant, with similar trends for known and unknown items. When tested separately, however, illusory truth occurred for unknown, but not known, statements. Ceiling effects in the known condition (i.e., strong bias to respond "true") render these data inconclusive.

To summarize, illusory truth generalizes across a remarkably wide range of factors. In the absence of source recollection, the only constraint commonly identified is that participants must lack knowledge about the statements' veracity. In two experiments, we evaluate the claim that illusory truth effects do not occur if people can draw upon their stored knowledge. We used two types of statements: contradictions of well-known facts and contradictions of facts unknown to participants. We defined knowledge using Nelson and Narens's (1980) norms, as well as

individuals' performance on a post-experimental knowledge check (Experiment 1). Additionally, we created two competing multinomial models of the way people evaluate statements' truthfulness. The knowledge-conditional model reflects the assumption that people search memory for relevant information, only relying on fluency if this search is unsuccessful. The fluency-conditional model, on the other hand, posits that people can rely solely on fluency, even if stored knowledge is available to them. We tested the fit of these models of illusory truth using binary data (Experiment 2).

Experiment 1

Method

Participants. Forty Duke University undergraduates participated in exchange for monetary compensation. Participants were tested individually or in small groups of up to five people.

Design. The experiment had a 2 (repetition: repeated, new) X 2 (estimated knowledge: known, unknown) within-subjects design. Both factors were counterbalanced across participants.

Materials. We selected 176 questions from Nelson and Narens's (1980) general knowledge norms, half of which were likely to be known (on average, answered correctly by 60% of norming participants) and half of which were likely to be unknown (answered correctly by only 5% of norming participants).[1] These norms likely underestimate how many facts participants can classify as true or false, since the norming study required participants to produce answers to open-ended questions. For each question, we created a truth (e.g. *Photosynthesis is the name of the process by which plants make their food.*) and a matching falsehood that referred to a plausible, but incorrect, alternative (e.g. *Chemosynthesis is the name of the process by which*

plants make their food.). This resulted in four item types: known truths, known falsehoods, unknown truths, and unknown falsehoods. To be clear, a "known falsehood" refers to a contradiction of a fact stored in memory; participants did not receive explicit labels or any other indications that specific statements were true or false. Sample statements can be seen in Table 1.

We divided both the known and unknown items into four sets of 22 statements. Two known and two unknown sets appeared as truths, and the remainder appeared as falsehoods; furthermore, half of the truths and falsehoods repeated across exposure and truth rating phases, whereas the other half appeared for the first time during the truth rating phase. Given our interest in how people evaluate false claims, we limited our analyses to falsehoods and treated truths as fillers. Additionally, responses to known truths averaged "probably true," leaving little room for repetition to bias judgments.

Procedure. After giving informed consent, participants completed the first phase of the experiment, the *exposure phase.* Participants rated 88 statements for subjective interest, using a 6-point scale labeled 1 = *very interesting*, 2 = *interesting*, 3 = *slightly interesting*, 5 = *slightly uninteresting*, 5 = *uninteresting*, and 6 = *very uninteresting*. The experimenter informed participants that their ratings would guide stimulus development for future experiments, and that some statements were true and others false.

Immediately after exposure, participants completed the second part of the experiment, the *truth rating phase*. In addition to the warning that they would encounter true and false statements, the experimenter told participants that some statements appeared earlier in the experiment, while others were new. Participants rated 176 statements for truthfulness, using a scale labeled 1 = *definitely false*, 2 = *probably false*, 3 = *possibly false*, 4 = *possibly true*, 5 = *probably true*, and 6 = *definitely true*.

The norms provide useful information about which questions are relatively easy or difficult for participants, as documented in numerous studies (Eslick, Fazio, & Marsh, 2011; Marsh & Fazio, 2006; Marsh, Meade, & Roediger, 2003). However, the norms cannot predict with perfect accuracy what each individual knows. To address this concern, we borrowed a procedure used in other experiments (e.g., Kamas, Reder, & Ayers, 1996). Following the exposure and truth rating phases, participants completed the *knowledge check phase*. They answered 176 multiple-choice questions with three response options: the correct answer, the alternative from the false version of each statement, and "don't know." For example, the answer options *Atlantic*, *Pacific*, and *don't know* accompanied the question *What is the largest ocean on Earth?*.

Results

The alpha level for all statistical tests was set to .05. As discussed above, we focused our analyses primarily on falsehoods.

Knowledge check. We first assessed knowledge check performance, to ensure adequate proportions of known and unknown items across participants. Overall, participants answered 44% of the knowledge check questions correctly (known items). They responded to 12% of the questions with falsifications, and to another 44% 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, the knowledge check underestimates people's knowledge, since viewing the false version of a statement may bias people to later choose the wrong answer (Bottoms, Eslick, & Marsh, 2010; Kamas et al., 1996).

Results of the knowledge check confirmed estimates of knowledge based on norms. Participants correctly answered 67% of the questions estimated by the norms to be known and 20% of those estimated to be unknown (compared to 60% and 5%, respectively, in the original norming study). Participants indicated "don't know" for 23% for the questions estimated to be known and for 65% of those estimated to be unknown, again suggesting that participants did not guess.

Truth ratings. We analyzed truth ratings as a function of both (a) individual knowledge check performance and (b) norm-based estimates of knowledge. To preview, these analyses returned very similar results.

Truth ratings as a function of demonstrated knowledge. We conducted a 2 (repetition: repeated, new) X 2 (demonstrated knowledge: known, unknown) repeated-measures ANOVA on participants' truth ratings of falsehoods. The number of known and unknown items varied for each participant, depending upon on his or her knowledge check performance. Every participant's data included a minimum of 5 trials per cell, and the average trial count per cell was 22. The relevant data appear in Figure 1A. Replicating the illusory truth effect, repeated falsehoods (M = 3.53) received higher truth ratings than new ones (M = 3.26), $F(1, 39) = 13.06$, $MSE = .23$, $p = .001$, $\eta_p^2 = .25$. As expected, known falsehoods (M = 2.76) received lower (i.e., more accurate) ratings during the truth phase than unknown ones (M = 4.03), $F(1, 39) = 196.05$, $MSE = .33$, $p < .001$, $\eta_p^2 = .83$.

Surprisingly, knowledge did not interact with repetition to influence truth ratings, $F(1, 39) = 1.21$, $p = .28$. Similar illusory truth effects occurred for contradictions of known (repeated M = 2.93; new M = 2.59; $t(39) = 2.93$, $SEM = .12$, $p = .006$) and unknown (repeated M = 4.13; new M = 3.92; $t(39) = 2.79$, $SEM = .07$, $p = .008$) statements. In fact, the effect was numerically

larger in the known condition. In other words, repetition increased perceived truthfulness, even for contradictions of well-known facts.

For completeness, we note that an illusory truth effect also emerged for truths (repeated $M = 4.11$; new $M = 3.87$; $t(39) = 3.41$, $SEM = .07$, $p = .002$). Illusory truth occurred for unknown (4.11 vs. 3.82; $t(39) = 3.28$, $SEM = .09$, $p = .002$) but not known (5.28 vs. 5.15; $t(39) = 1.80$, $SEM = .07$, $p = .08$) statements. The latter result likely reflects a ceiling effect.

Truth ratings as a function of knowledge estimated by norms. We also analyzed the data using norm-based estimates of knowledge. The complete data appear in Figure 1B, but we will highlight only the key result: Participants exhibited illusory truth effects for both known (repeated $M = 3.41$; new $M = 2.94$; $t(39) = 5.02$, $SEM = .09$, $p < .001$) and unknown (repeated $M = 3.81$; new $M = 3.53$; $t(39) = 3.41$, $SEM = .08$, $p < .001$) falsehoods.

Discussion

Experiment 1 tested the widely-held assumption that illusory truth effects depend upon the absence of knowledge. Surprisingly, repetition increased statements' perceived truth, regardless of whether stored knowledge could have been used to detect a contradiction. Reading a statement like *A sari is the name of the short pleated skirt worn by Scots* increased participants' later belief that it was true, even if they could correctly answer the question *What is the name of the short pleated skirt worn by Scots?*. Similar results were observed when knowledge was estimated using group norms or directly measured in individuals using a post-experiment knowledge check. We also replicated this pattern using norm-based estimates of knowledge in a smaller experiment ($n = 16$) that we will not report in detail, other than to note that the interaction between knowledge and repetition was not significant, $F(1, 15) = 1.43$, $MSE = .07$, p

= .25, and that there was a robust illusory truth effect for known falsehoods, $t(15) = 3.57$, $SEM = .13$, $p = .003$.

These data suggest a counterintuitive relationship between fluency and knowledge. Prior work assumes that people only rely on fluency if knowledge retrieval is unsuccessful (i.e., if participants lack relevant knowledge or fail to search memory at all). Experiment 1 demonstrated that the reverse may be true: Perhaps people retrieve their knowledge only if fluency is absent (i.e., if statement is new or was not attended to during the exposure phase; if the participant spontaneously discounts fluency while reading repeated statements). To discriminate between these two possibilities, we created multinomial models in the form of branching tree diagrams with parameters representing unobserved cognitive processes. Each parameter represents the probability that the cognitive process contributes to the observed behavior (from 0 to 1). This method has successfully characterized diverse phenomena, including the hindsight bias (Erdfelder & Buchner, 1998), the misinformation effect (Jacoby, Bishara, Hessels, & Toth, 2005), and the engagement of racial stereotypes (Bishara & Payne, 2009).

The *knowledge-conditional model* assumes that when judging a statement's truthfulness, people search their memory for relevant evidence (see Figure 2). If this search succeeds (probability = K), all other processes are irrelevant, and the participant answers correctly. If the search fails (probability = 1 – K), due to a lack of knowledge or insufficient cues, the participant may rely on fluency to make the judgment. If the participant relies on fluency (probability = F), he or she exhibits a bias to respond "true"; if fluency is absent or discounted (probability = 1 – F), the participant guesses "true" (probability = G) or "false" (probability = 1 – G).

In contrast, the *fluency-conditional model* utilizes a different set of conditional probabilities in assuming that fluency can supersede retrieval of knowledge (see Figure 3). The

participant only searches memory if fluency is absent or discounted (probability = 1 – F). In this case, the participant retrieves knowledge (probability = K) or if nothing is retrieved (probability = 1-K), he or she guesses "true" (probability = G) or "false" (probability = 1 – G). In those cases where the participant experiences fluent processing and does not discount that fluency, the participant exhibits a bias to respond "true" (probability = F).

At first glance, these models may appear to be formally equivalent. While they contain the same parameters, the graphical order (and thus conditional probabilities) of the models differ markedly. To better conceptualize how the models differ, it is useful to remember Bayes' Theorem, whereby the probability of A given B is not necessarily equal to the probability of B given A. For example, the probability that a person is male, given that he/she is 45 years old, is not equal to the probability that a person is 45 years old, given that he is male. Applied to the present models, failing to retrieve knowledge given disfluent processing is not necessarily equivalent to the probability of disfluent processing given failure to retrieve knowledge. Figures 2 and 3 illustrate how these two models lead to different predicted responses.

To facilitate model testing, we replicated Experiment 1 with a binary truth rating rather than a 6-point scale. The structure of the models requires that we analyze both truths and falsehoods; if we exclusively modeled falsehoods, we would be unable to discriminate knowledge retrieval from accurate guesses. Because Experiment 1 validated the use of norms in estimating participants' knowledge, this experiment did not employ a knowledge check. We used these binary data to compare the relative fits of the knowledge-conditional and fluency-conditional models of illusory truth.

Experiment 2

Method

Participants. Forty Duke University undergraduates participated in exchange for monetary compensation. Participants were tested individually or in small groups of up to five people.

Design. The experiment had a 2 (truth status: truth, falsehood) X 2 (repetition: repeated, new) X 2 (estimated knowledge: known, unknown) within-subjects design. All factors were counterbalanced across participants.

Materials. We used the same statements in Experiments 1.

Procedure. The procedure was identical to that of Experiment 1, with the exceptions that (a) participants made binary truth judgments (*true* or *not true*) instead of using a 6-point scale and (b) there was no knowledge check.

Results

Unlike Experiment 1, we performed analyses on the proportion of statements rated "true."

Truth ratings. The illusory truth effects in Experiment 1 represent small shifts along the middle of a 6-point scale. To confirm that we replicated Experiment 1 with a binary scale, we conducted a 2 (truth status: truth, falsehood) X 2 (repetition: repeated, new) X 2 (estimated knowledge: known, unknown) ANOVA on the proportion of statements judged to be "true." The complete data appear in Figure 4; as expected there were main effects of truth status, $F(1, 39) = 163.81$, $MSE = .03$, $p < .001$, $\eta_p^2 = .81$, and estimated knowledge, $F(1, 39) = 25.65$, $MSE = .04$, $p < .001$, $\eta_p^2 = .40$. In addition, the basic illusory truth effect emerged: Repeated statements ($M = 0.62$) were more likely to be judged "true" than new statements ($M = 0.56$), $F(1, 39) = 13.18$,

MSE = .03, p = .001, η_p^2 = .25. Critically, there was no interaction between repetition and estimated knowledge, $F(1, 39) = 2.36$, p = .13, *MSE* = .01; illusory truth occurred regardless of whether the statements were estimated to be known (.67 vs. .62, $t(39) = 2.34$, *SEM* = .02, p = .025) or unknown (.58 vs. .49, $t(39) = 3.53$, *SEM* = .02, p = .001). We observed a significant three-way interaction among truth, repetition, and knowledge, reflecting a ceiling effect for known truths, $F(1, 39) = 6.86$, *MSE* = .01, p = .01, η^2 = .15.

Model testing. We tested the fit of both multinomial models using multiTree software (Moshagen, 2010), which minimizes a G^2 statistic (see Appendix for equations); lower G^2 values indicate better model fit. The null hypothesis states that the model fits, so significant p values indicate poor fit. To preserve degrees of freedom, we placed theoretically-motivated constraints on the parameters. These constraints ensured that there were more data cells (8) than free parameters (5) in each model. Specifically, the knowledge parameter (K) was free to vary across known and unknown statements, but was constrained to be equal for truths and falsehoods, as well as for new and repeated statements. Second, the fluency parameter (F) was free to vary across new and repeated statements, but was constrained to be equal for known and unknown statements, as well as for truths and falsehoods. Finally, the guessing parameter (G) was held constant across all cells.

With these constraints, the knowledge-conditional model fit the data poorly, G^2 (df = 3) = 185.59, $p < .00001$. In contrast, the fluency-conditional model fit well, G^2 (df = 3) = 2.54, p = 0.47. Furthermore, as shown in Table 2, the parameter estimates for the fluency-conditional model reflect the main effects reported in Experiment 1. Knowledge retrieval occurred more for statements estimated to be known (K = .72) rather than unknown (K = .13), reflecting the main effect of knowledge. In addition, reliance on fluency was higher for repeated (F = .44) than new

statements (F = .35), reflecting the main effect of repetition. The one result that may appear counterintuitive is the relatively low probability of guessing "true" (G = .21). However, this estimate is likely less accurate than the estimates of the other parameters, given the wide confidence interval surrounding it [0.13, 0.28]. In addition, guessing only influenced responses in a select few cases, where fluency was absent or discounted *and* knowledge retrieval failed. In other words, any influence of guessing was likely small.

We also compared the fit of the fluency-conditional model to a nested model with additional parameter constraints. Here, the null hypothesis states that the nested model fits as well as the original model. Constraining K to be equal for known and unknown statements led to poor model fit, ΔG^2 ($\Delta df = 1$) = 380.64, $p < .00001$, as did constraining F to be equal for new and repeated statements, ΔG^2 ($\Delta df = 1$) = 30.78, $p < .001$. Finally, we tested a version of the model where F was not constrained to be equivalent for true and false statements (based on the arguments in Unkelbach, 2007). This modified model (which allowed F to vary across truths and falsehoods) had too few free parameters, but the parameter values varied as expected.

It is important to note that the fluency-conditional model of illusory truth is entirely consistent with the large main effects of knowledge reported in Experiments 1 ($\eta_p^2 = .83$) and 2 ($\eta_p^2 = .40$). The superior fit of the fluency-conditional model demonstrates that it is *possible* for participants to rely on fluency despite having contradictory knowledge stored in memory (an impossibility in the knowledge-conditional model). This model does not assume that participants will always rely on fluency; in fact, estimates of fluency-based responding were relatively low (F < .5 across all trials). Instead, participants relied on knowledge if statements lacked fluency (i.e., they were new or were not well-attended while rating interest) or because participants spontaneously discounted their feelings of fluency upon reading some of the repeated statements

(Oppenheimer, 2004). Under either of these circumstances, participants searched memory for relevant evidence, yielding a main effect of knowledge. In other words, the question of which process is conditional on the other is separate from whether the probability of either process is high or low.

To summarize, the results of our model testing complement the findings in Experiment 1. To further validate the modeling, we dichotomized the truth judgments from Experiment 1 and assessed which model fit those data. When we estimated knowledge with the norms, the fluency-conditional model fit the data well, G^2 (df = 3) = 2.10, p = 0.55, while the knowledge-conditional model did not, G^2 (df = 3) = 63.44, p < .00001. Similar patterns held when we defined knowledge at at an individual level (i.e., performance on the knowledge check): The fluency-conditional model fit well, G^2 (df = 3) = 6.22, p = 0.10, while the knowledge-conditional model fit poorly, G^2 (df = 3) = 65.43, p < .00001. Thus, re-analysis of Experiment 1 data yielded the same conclusions; neither study supported the knowledge-conditional model, which has been assumed in the literature until now. The fluency-conditional model, on the other hand, fit the data well. People searched memory in the absence of fluency, consistent with the idea that disfluent processing triggers more elaborate processing (Song & Schwartz, 2008).

General Discussion

The present research demonstrates that fluency can influence people's judgments, even in contexts that allow them to draw upon their stored knowledge. The results of two experiments suggest that people sometimes fail to bring their knowledge to bear and instead rely on fluency as a proximal cue. Participants more accurately judged the truth of known than unknown statements, but there was no interaction between knowledge and repetition. Whether we defined knowledge using norm-based estimates or individuals' accuracy on a knowledge check, fluency

exerted a similar effect on contradictions of well-known and ambiguous facts. Our conclusions do not contradict the few studies targeting the moderating role of knowledge in illusory truth. As noted earlier, the data on expertise do not really speak to the issue at hand, as those studies targeted ambiguous statements within an expert domain (Arkes et al., 1989; Boehm, 1994). The data also do not contradict Unkelbach and Stahl's (2009) multinomial model, which included a knowledge parameter intentionally set to be near zero. The fluency-conditional model extends their work to situations in which people evaluate information already stored in memory.

While our findings contradict a dominant assumption, they are consistent with what we know about semantic retrieval, where the knowledge retrieved often lacks source information (Tulving, 1972). While people can recall and evaluate source information when judging recently acquired information (represented by Unkelbach & Stahl's (2009) recollection parameter), people rarely engage in source monitoring when evaluating information stored in their knowledge bases. This non-evaluative tendency may render people especially susceptible to external influences like fluency. Kelley and Lindsay (1993), for example, demonstrated the influence of *retrieval fluency*, or the ease with which an answer is retrieved from memory. Participants read a series of words, some of which were semantically related to the answers for a later general knowledge test. Later, participants not only incorrectly answered questions with the lures they saw earlier, but did so with high confidence, in what the authors termed "illusions of knowledge." These data bear a strong resemblance to the present findings, where people underutilized their knowledge in the face of repetition-based fluency.

Knowledge neglect, or the failure to appropriately apply stored knowledge, occurs in tasks other than truth judgments. Fazio and Marsh (2008), for example, exposed participants to errors embedded in fictional stories (e.g., *paddling across the largest ocean, the Atlantic*). Errors

contradicted known or unknown facts, in a knowledge manipulation similar to that of Experiment 1. Participants were no better at detecting contradictions with known than unknown facts. Similarly, in the *Moses illusion*, participants answer a series of questions, some of which include faulty presuppositions (e.g., *How many animals of each kind did Moses take on the ark?*). People often answer this question as if nothing were wrong with it, despite knowing that Noah, not Moses, took animals on the ark (Erickson & Mattson, 1981). The present data reveal that knowledge neglect occurs even when participants explicitly evaluate statements' truthfulness.

In the experiments reported here, participants sometimes neglected their knowledge under fluent processing conditions. Gilbert (1991) argues that people automatically assume that a statement is true because "unbelieving" comprises a second, resource-demanding step. Even when people devote resources to evaluating a claim, they only require a "partial match" between the contents of the statement and what is stored in memory (see Reder & Cleeremans, 1990; Reder & Kusbit, 1991). In other words, we tend to notice errors that are less semantically related to the truth (e.g., to notice the error in the question *How many animals of each kind did Adam take on the ark?*; van Oostendorp & de Mul, 1990). We expect that participants would draw on their knowledge, regardless of fluency, if statements contained implausible errors (e.g., *A grapefruit is a dried plum*, instead of *A date is a dried plum*); in this example, the limited semantic overlap of the words "grapefruit" and "plum" would yield an insufficient match. Critically, what constitutes a partial match depends on an individual's knowledge. Experts may be less susceptible to illusory truth, as long as they possess knowledge directly pertinent to the statements (unlike Boehm, 1994, and Arkes et al., 1989).

Contrary to the connotations of the term "illusory truth" and "knowledge neglect," fluency serves as a useful cue in many everyday situations. Inferring truth from fluency often proves to be an accurate and cognitively inexpensive strategy, making it reasonable that people sometimes apply this heuristic without searching for knowledge. However, certain situations likely discourage the use of the fluency heuristic; fact-checkers reading drafts of a magazine article or reporters waiting to catch a politician in a misstatement, for example, likely draw on knowledge in the face of fluency. In addition, sufficient experience may encourage an individual to shift from a fluency-conditional to a knowledge-conditional approach. As an example, learners provided with trial-by-trial feedback may learn that their gut responses are often wrong. Our work demonstrates that fluency can emerge as the dominant signal in some contexts, but future research should examine the factors that encourage reliance on knowledge instead.

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. doi:10.1177/1088868309341564

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*, 576-605. doi: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*, 81-94. doi:10.1002/bdm.3960020203

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

Barber, S. J., Rajaram, S., & Marsh, E. J. (2008). Fact learning: How information accuracy, delay, and repeat testing change retention and retrieval experience. *Memory*, *16*, 934-936. doi:10.1080/09658210802360603

Begg, I., 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*, 446-458. doi:10.1037/0096-3445.121.4.446

Begg, I., & Armour, V. (1991). Repetition and the ring of truth: Biasing comments. *Canadian Journal of Behavioural Science*, *23*, 195-213. doi:10.1037/h0079004

Begg, I., Armour, V., & Kerr, T. (1985). On believing what we remember. *Canadian Journal of Behavioural Science*, *17*, 199-214. doi:10.1037/h0080140

Bishara, A., & Payne, B. K. (2009). Multinomial process tree models of control and automaticity in weapon misidentification. *Journal of Experimental Social Psychology*, *45*, 524-534. doi:10.1016/j.jesp.2008.11.002

Boehm, L. E. (1994). The validity effect: A search for mediating variables. *Personality and Social Psychology Bulletin*, *20*, 285-293. doi: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. doi:10.1080/09658211.2010.501558

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. doi:10.1037/0278-7393.22.5.1088

Conway, M. A., Gardiner, J. M., Perfect, T. J., Anderson, S. J., & Cohen, G. M. (1997). Changes in memory awareness during learning: The acquisition of knowledge by psychology undergraduates. *Journal of Experimental Psychology: General*, *126*, 393-413. doi:10.1037/0096-3445.126.4.393

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

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. doi:10.1177/1088868309352251

Dewhurst, S. A., Conway, M. A., & Brandt, K. R. (2009). Tracking the R-to-K shift: Changes in memory awareness across repeated tests. *Applied Cognitive Psychology*, *23*, 849-858. doi:10.1002/acp.1517

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 387-414. doi:10.1037/0278-7393.24.2.387

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning*, *20*, 540-551. doi:10.1016/S0022-5371(81)90165-1

Eslick, A. N., Fazio, L. K., & Marsh, E. J. (2011). Ironic effects of drawing attention to story errors. *Memory*, *19*, 184-191. doi:10.1080/09658211.2010.543908

Fazio, L. K., & Marsh, E. J. (2008). Slowing presentation speed increases illusions of knowledge. *Psychonomic Bulletin & Review*, *15*, 180-185. doi:10.3758/PBR.15.1.180

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology*, *97*, 185-195. Retrieved from http://www.press.uillinois.edu/journals/ajp.html

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119. doi:10.1037/0003-066X.46.2.107

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning & Verbal Behavior*, *16*, 107-112. doi:10.1016/ S0022-5371(77)80012-1

Hawkins, S. A., & Hoch, S. J. (1992). Low-involvement learning: Memory without evaluation. *Journal of Consumer Research*, *19*, 212-225. doi:10.1086/209297

Henkel, L. A., & Mattson, M. E. (2011). Reading is believing: The truth effect and source credibility. *Consciousness and Cognition*, *20*, 1705-1721. doi:10.1016/j.concog.2011.08.018

Herbert, D. M. B., & Burt, J. S. (2001). Memory awareness and schematization: Learning in the university context. *Applied Cognitive Psychology*, *15*, 617-637. doi:10.1002/acp.729

Herbert, D. M. B., & Burt, J. S. (2003). The effects of different review opportunities on schematisation of knowledge. *Learning and Instruction*, *13*, 73-92. doi:10.1016/S0959-4752(01)00038-X

Herbert, D. M. B., & Burt, J. S. (2004). What do students remember? Episodic memory and the development of schematization. *Applied Cognitive Psychology*, *18*, 77-88. doi:10.1002/acp.947

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995-1006. doi:10.1037/0022-3514.79.6.995

Jacoby, L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*, *134*, 131-148. doi:10.1037/0096-3445.134.2.131

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. doi: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. doi:10.3758/BF03201094

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, *32*, 1-24. doi:10.1006/jmla.1993.1001

Marsh, R. L., & Bower, G. H. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Unconscious plagiarism in a puzzle task*, *19*, 673-688. doi:10.1037/0278-7393.19.3.673

Marsh, E. J., & Fazio, L. K. (2006). Learning errors from fiction: Difficulties in reducing reliance on fictional stories. *Memory & Cognition*, *34*, 1140-1149. doi:10.3758/BF03193260

Marsh, E. J., Meade, M. L., & Roediger, H. L. (2003). Learning facts from fiction. *Journal of Memory and Language*, *49*, 519-536. doi:10.1016/S0749-596X(03)00092-5

Marsh, R. L., Landau, J. D., & Hicks, J. L. (1997). Contributions of inadequate source monitoring to unconscious plagiarism during idea generation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 886-897. doi:10.1037/0278-7393.23.4.886

Merritt, P., Hirshman, E., Zamani, S., Hsu, J., & Berrigan, M. (2006). Episodic representations support early semantic learning: Evidence from midazolam induced amnesia. *Brain and Cognition*, *61*, 219-223. doi:10.1016/j.bandc.2005.12.001

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*, 42-54. doi:10.3758/BRM.42.1.42

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning & Verbal Behavior*, *19*, 338-368. doi:10.1016/S0022-5371(80)90266-2

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

Parks, C. M., & Toth, J. P. (2006). Fluency, familiarity, and the illusion of truth. *Aging, Neuropsychology, and Cognition*, *13*, 225-253. doi:10.1080/138255890968691

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

Reder, L. M., & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In A. Graesser & G. Bower (Eds.), *The Psychology of Learning and Motivation* (Vol. 25, pp. 233–258). New York: Academic Press.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language, 30*, 385-406. doi:10.1016/0749-596X(91)90013-A

Schwartz, M. (1982). Repetition and rated truth value of statements. *The American Journal of Psychology*, *95*, 393-407. doi:10.2307/1422132

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18*, 1074-1083. doi:10.1037/0278-7393.18.5.1074

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

Srull, T. K. (1983). The role of prior knowledge in the acquisition, retention, and use of new information. *Advances in Consumer Research*, *10*, 572-576. Retrieved from http://www.acrwebsite.org/

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. doi:10.3758/s13428-012-0307-9

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381-402). New York, NY: Academic Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232. doi:10.1016/0010-0285(73)90033-9

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. doi:10.1037/0278-7393.33.1.219

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

van Oostendorp, H., & de Mul, S. (1990). Moses beats Adam: A semantic relatedness effect on a semantic illusion. *Acta Psychologica*, *74,* 35-46. doi:10.1016/0001-6918(90)90033-C

Footnotes

[1] At the time of data collection, Tauber, Dunlosky, Rawson, Rhodes, and Sitzman's (2013) updated norms were not yet published.

Appendix: Modeling Methods

Multinomial models were implemented using multiTree software (Moshagen, 2010) with random start values. With an alpha level of .05, power to detect medium effect sizes (w = .3; Cohen, 1977) always exceeded .999.

Knowledge-Conditional Model Equations:

$$P(\text{Correct}|\text{True}) = K + (1 - K)(F) + (1 - K)(1 - F)(G) \quad \text{(A1)}$$

$$P(\text{Incorrect}|\text{True}) = (1 - K)(1 - F)(1 - G) \quad \text{(A2)}$$

$$P(\text{Correct}|\text{False}) = K + (1 - K)(1 - F)(1 - G) \quad \text{(A3)}$$

$$P(\text{Incorrect}|\text{False}) = (1 - K)(F) + (1 - K)(1 - F)(G) \quad \text{(A4)}$$

Fluency-Conditional Model Equations:

$$P(\text{Correct}|\text{True}) = F + (1 - F)(K) + (1 - F)(1 - K)(G) \quad \text{(A5)}$$

$$P(\text{Incorrect}|\text{True}) = (1 - F)(1 - K)(1 - G) \quad \text{(A6)}$$

$$P(\text{Correct}|\text{False}) = (1 - F)(K) + (1 - F)(1 - K)(1 - G) \quad \text{(A7)}$$

$$P(\text{Incorrect}|\text{False}) = F + (1 - F)(1 - K)(G) \quad \text{(A8)}$$

Table 1

Sample Known and Unknown Statements

	Truth	Falsehood
Known	A prune is a dried plum.	A date is a dried plum.
	The Pacific Ocean is the largest ocean on Earth.	The Atlantic Ocean is the largest ocean on Earth.
	The Cyclops is the legendary one-eyed giant in Greek mythology.	The Minotaur is the legendary one-eyed giant in Greek mythology.
Unknown	Helsinki is the capital of Finland.	Oslo is the capital of Finland.
	Marconi is the inventor of the wireless radio.	Bell is the inventor of the wireless radio.
	Billy the Kid's last name is Bonney.	Billy the Kid's last name is Garrett.

Table 2

Parameter Estimates for the Fluency-Conditional Model

	Known		Unknown	
Parameter	New	Repeated	New	Repeated
F	0.35 [0.31, 0.39]	0.44 [0.40, 0.48]	0.35 [0.31, 0.39]	0.44 [0.40, 0.48]
K	0.72 [0.69, 0.76]	0.72 [0.69, 0.76]	0.13 [0.07, 0.19]	0.13 [0.07, 0.19]
G	0.21 [0.13, 0.28]	0.21 [0.13, 0.28]	0.21 [0.13, 0.28]	0.21 [0.13, 0.28]

Note. Parameter values reflect the probability that the cognitive processes contribute to the observed behavior (from 0 to 1). The 95% confidence interval around each parameter estimate is noted in brackets. F = reliance on fluency; K = retrieval of knowledge; G = guess "true."

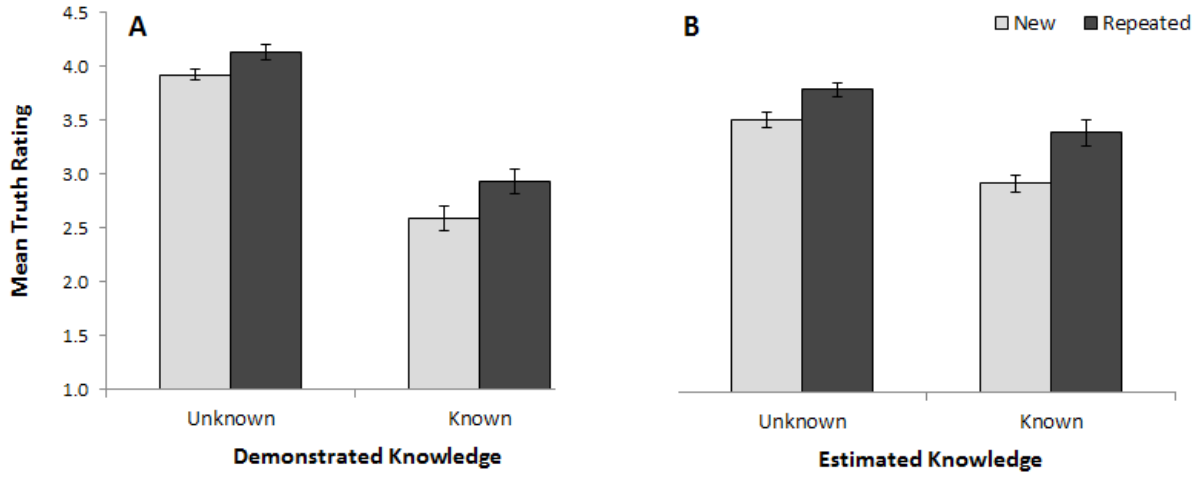


Figure 1. Mean truth ratings for falsehoods as a function of repetition and both demonstrated (A) and normed-based estimates of (B) knowledge (Experiment 1). Error bars reflect standard error of the mean.

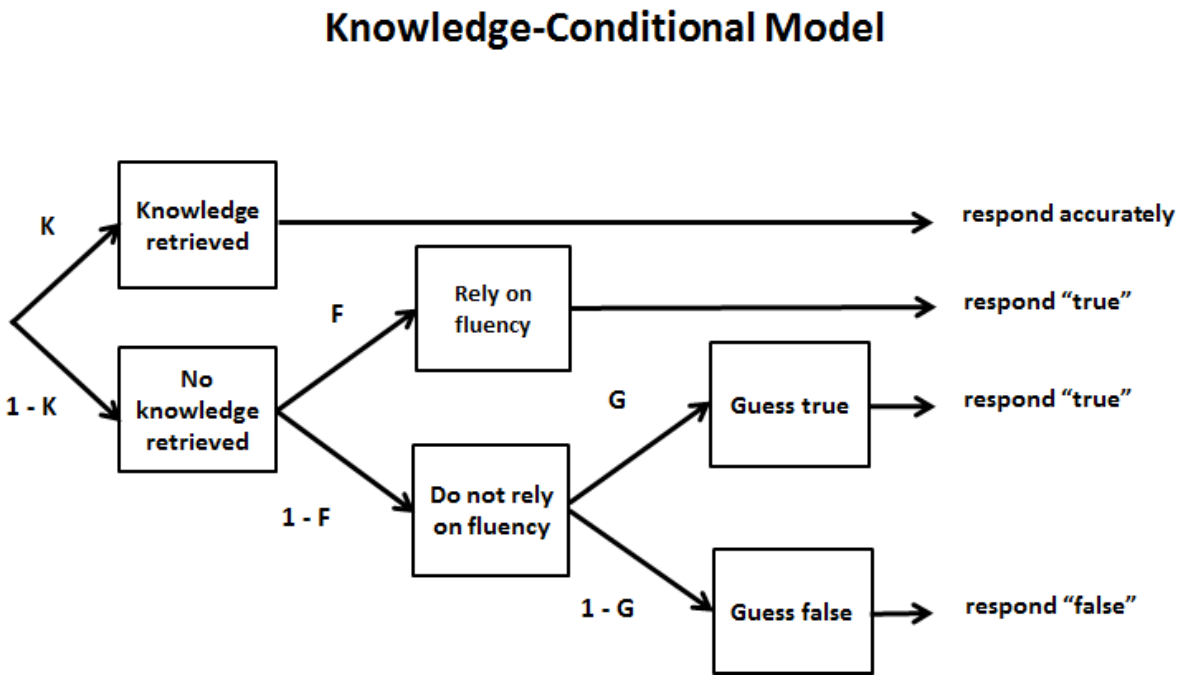


Figure 2. Knowledge-conditional model of illusory truth. Parameter values reflect the probability that the cognitive process contributes to behavior (from 0 to 1). K = knowledge; F = fluency; G = guess true.

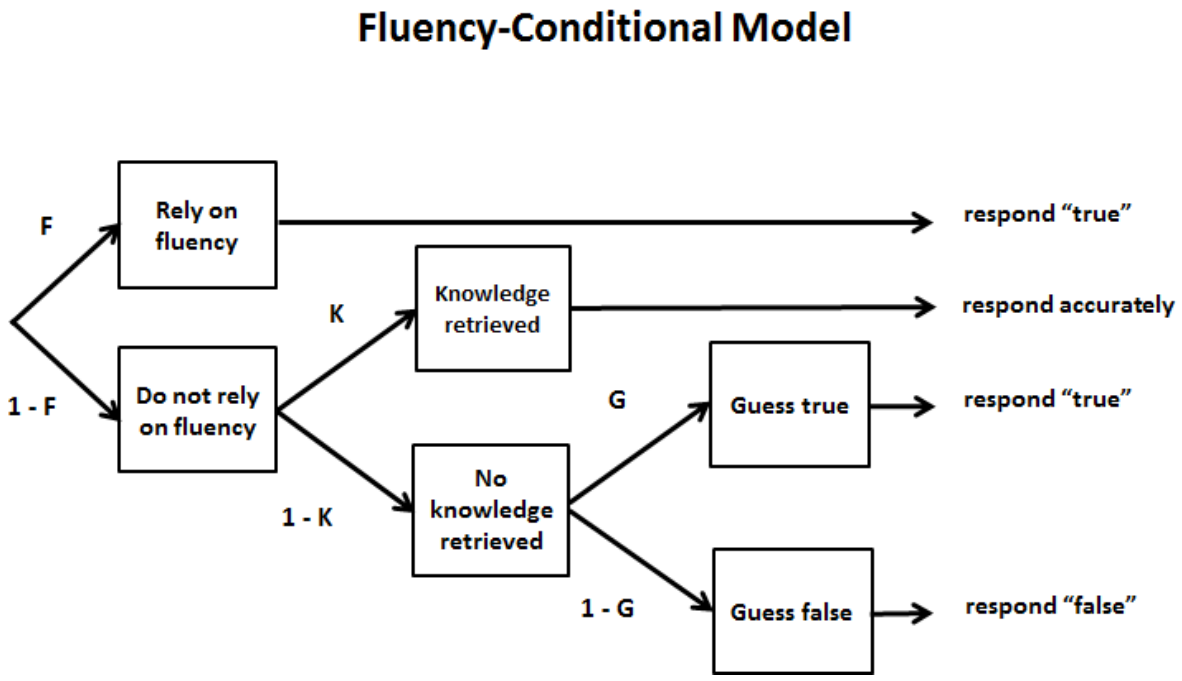


Figure 3. Fluency-conditional model of illusory truth. Parameter values reflect the probability that the cognitive process contributes to behavior (from 0 to 1). K = knowledge; F = fluency; G = guess true.

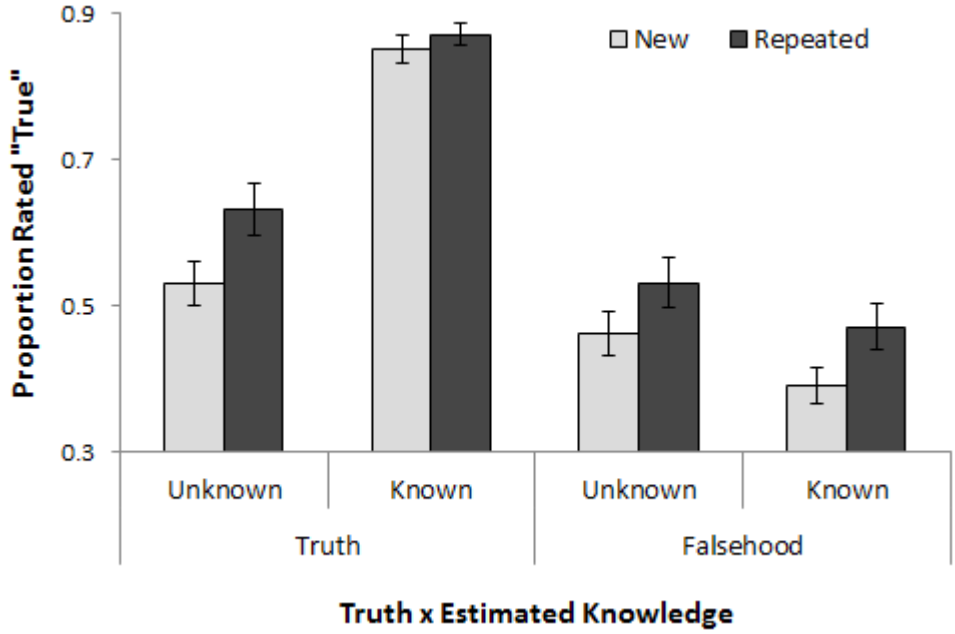


Figure 4. Proportion of statements rated "true" as a function of repetition, truth, and norm-based estimates of knowledge (Experiment 2). Error bars reflect standard error of the mean.