Acta Psychologica 179 (2017) 114–123

Contents lists available at ScienceDirect

Acta Psychologica

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

The citation effect: In-text citations moderately increase belief in trivia claims

Adam L. Putnam*, Riley J. Phelps

Carleton College, United States

ARTICLE INFO

Keywords:
Truth judgments
Cognitive fluency
Decision making
Memory
Truthiness

ABSTRACT

Authors use in-text citations to provide support for their claims and to acknowledge work done by others. How much do such citations increase the believability of an author's claims? It is possible that readers (especially novices) might ignore citations as they read. Alternatively, citations ostensibly serve as evidence for a claim, which justifies using them as a basis for a judgment of truth. In six experiments, subjects saw true and false trivia claims of varying difficulty presented with and without in-text citations (e.g., *The cat is the only pet not mentioned in the bible*) and rated the likelihood that each statement was true. A mini meta-analysis summarizing the results of all six experiments indicated that citations had a small but reliable effect on judgments of truth ($d = 0.13$, 95% CI [0.06, 0.20]) suggesting that subjects were more likely to believe claims that were presented with citations than without. We discuss this *citation effect* and how it is similar and different to related research suggesting that nonprobative photos can increase judgments of truth.

1. Introduction

One foundation of good critical thinking is the ability to evaluate the credibility of claims. "Can sharks swim backwards?" "Was President Obama born in the United States?" "Can vaccines cause autism?" In today's society we are inundated with facts, stories, and claims from myriad online sources that vary widely in credibility. After a tumultuous presidential election, concerns about truth and credibility have become a national issue in 2017 – Time magazine even featured a cover story with the title "Is Truth Dead?" (Scherer, 2017). Fortunately, a goal for many educators is to help their students become better critical thinkers; education should help students learn to critically read texts, skillfully evaluate evidence, and develop habits of skepticism. Such education is important, because research suggests that people accept statements or claims as true unless they are prompted in some way to look deeper into a claim's evidence, believability, or importance (see Gilbert, Krull, & Malone, 1990 for evidence; Gilbert, 1991 for an overview).[1]

In academic writing, one signal of evidence is the use of in-text or parenthetical citations (e.g., the Gilbert references above). Authors use references to acknowledge the work of others and to provide support for their claims. The current experiments were designed to examine how much parenthetical citations affect the believability of trivia claims. On one hand, in-text citations provide useful information. They show that authors have done their research and guide interested readers to external sources that will provide evidence. Thus, in-text citations are a probative source of information; it is rational to use them when forming a judgment of truth. On the other hand, it is unclear how much readers attend to in-text citations. Research suggests that non-experts (aka students) vary widely in the degree to which they look at citations while reading and how they use the information in a citation to draw inferences from the text (Sparks & Rapp, 2011; Strømsø, Bråten, Britt, & Ferguson, 2013).

Complicating matters further, several lines of research have demonstrated that truth judgments can be affected by a variety of factors, many of which are illogical or nonprobative (meaning they do not actually provide any additional diagnostic information). Newman, Garry, Bernstein, Kantner, and Lindsay (2012) use the term "truthiness" (borrowed from the comedian Stephen Colbert) to describe subjective feelings of truth. For example, in one widely read study, McCabe and Castel (2008) had students read science articles that included either pictures of a brain scan, a bar graph, or no accompanying image. The students who saw the brain images while reading rated the passage as having better scientific reasoning compared to students in the other conditions, even though the passages were identical. McCabe and Castel argued that the brain images were persuasive because they provided a

* Corresponding author at: Department of Psychology, Carleton College, 1 North College St, Northfield, MN 55057, United States.
E-mail address: aputnam@carleton.edu (A.L. Putnam).

[1] In case you were wondering, sharks cannot swim backwards, President Obama was born in Hawaii, U.S.A., and the link between vaccines and autism has been discredited (Colgrove & Bayer, 2005).

http://dx.doi.org/10.1016/j.actpsy.2017.07.010
Received 9 September 2016; Received in revised form 2 July 2017; Accepted 18 July 2017
Available online 04 August 2017

physical representation of an abstract cognitive idea. Although this brain image finding has been difficult to replicate (see Michael, Newman, Vuorre, Cumming, & Garry, 2013 for a meta-analysis), it does colorfully demonstrate how irrelevant information might affect someone's judgment.

A second line of research concerns what is called the *truth effect* (or sometimes the *illusory-truth effect*)–the finding that people are more likely to think that a statement is true if they have seen it before than if they are seeing it for the first time (Begg, Anas, & Farinacci, 1992; Hasher, Goldstein, & Toppino, 1977; for a meta-analysis see Dechêne, Stahl, Hansen, & Wänke, 2010). In other words, simply repeating a fact multiple times makes people more likely to believe it. Begg et al. (1992) and others (e.g., Nadarevic & Erdfelder, 2014; Unkelbach, 2007) have suggested that repeated statements are more familiar and that familiarity translates into more fluent processing of the item. The increased fluency is then mistakenly interpreted as a signal of truth. Begg et al. (1992) argued that it is illogical to use repetition in forming a judgment of truth (Unkelbach & Stahl, 2009 cite Wittgenstein as suggesting that using repetition to determine truth is like buying a second copy of a newspaper to see if the first is correct). In contrast, Unkelbach (2007; Reber & Unkelbach, 2010) has argued that repetition may be a valid basis for making a truth judgment; hearing a statement a second time–especially if it comes from a new source–provides converging evidence that the statement is true. Regardless of whether using repetition as a basis for truth is valid, it is clear that simply repeating a statement can increase the degree to which a statement is seen as true.

Moreover, fluency has been shown to influence truth judgments in a variety of ways, not just through increased familiarity. For example, people are more likely to believe that a trivia statement is true if it is presented in an easy to read format–a dark blue font against a white background–than a difficult to read format–a yellow font against a white background (Reber & Schwarz, 1999). Schwarz (2015) has argued that when people are deciding whether a claim is true they evaluate it against a set of five criteria (is the belief shared by others; is the belief supported by evidence; is the belief compatible with other things that one believes; does the belief have internal coherence; and is the source of the claim credible). Importantly, while people will evaluate different types of information for each of those criteria, fluency can affect the conclusions drawn from all of them. Information that is presented in an easy to process manner can inflate truth ratings via any of the above mechanisms.

One final striking example of how fluency can inflate truth ratings comes from a recent line of research that shows that presenting a photo along with a trivia claim makes subjects more likely to believe a statement, even when the photo does not provide any diagnostic information about the veracity of the claim (Cardwell, Henkel, Garry, Newman, & Foster, 2016; Fenn, Newman, Pezdek, & Garry, 2013; Newman et al., 2012; Newman et al., 2015). In one study (Newman et al., 2012, Experiment 3), for example, subjects saw a series of true and false trivia claims presented with or without an accompanying photo and were asked to judge whether the statements were true or false. Critically, all the photos were nonprobative – they were topically related to the claims, but did not provide any additional evidence about the truth of the claim. For example, the claim "Macadamia nuts are in the same evolutionary family as peaches" would appear with a picture of macadamia nuts. Despite not providing additional useful information, the photos led to a truth bias–subjects were more likely to accept a statement as true if it was presented with a photo than without. Newman et al. suggested that the photos helped people create "pseudoevidence"–subjects attributed the fluency of processing the photo as an indicator of truth, or used ambiguous information in the photo to confirm a hypothesis. Additional studies have shown that this truth bias persists over time (Fenn et al., 2013), that the photo has to be topically related to the trivia claim (e.g., the picture can't be completely unrelated, Newman et al., 2015), and that the presence of photos can even lead people to falsely remember past experiences (Cardwell et al., 2016).

In sum, truth judgments can be influenced by many factors, including ones that are nonprobative or irrational. Despite the research described above, no studies (to our knowledge) have examined whether in-text citations increase the perceived truthfulness of statements. Our interest in this question was partially inspired by an anonymous reviewer from a different paper (Putnam, Sungkhasettee, & Roediger, 2016) who suggested that including more references in a review on effective study strategies would make students more likely to believe the claims we made in our paper. We were skeptical that undergraduates would be persuaded by additional in-text citations and decided to investigate the question ourselves.

In the current experiments subjects saw true and false trivia statements presented with or without parenthetical citations and judged the truth of each statement. Across the experiments we used materials of varying difficulty, provided different instructions that sometimes emphasized what an in-text citation was, and manipulated the presence of citations both within and between subjects. Finally, we combined the evidence from each experiment in a mini meta-analysis to provide a more precise estimate of the effects of citations on truth judgments. Overall, we had two competing predictions. In contrast to the nonprobative photos used by Newman et al. (2012), citations are probative–they provide evidence or support for a claim. If nonprobative information can increase truth ratings, then probative information should as well. Therefore, our first hypothesis was that presenting in-text citations would increase the perceived truthfulness of the statements. Alternatively, parenthetical citations lack the visual appeal of photos and readers might ignore citations unless prompted to examine them (e.g., Gilbert, 1991). Thus, our second hypothesis was that subjects would provide similar truth ratings for statements presented with and without a citation.

2. Experiment 1A

Experiments 1A and 1B were identical, except that the variable *citations* was manipulated between-subjects in 1A and within-subjects in 1B. We expected that highlighting the difference between a statement with a citation and a statement without a citation (i.e., using a within-subjects design) would be more likely to show that citations affected truth ratings, whereas the between-subjects design would provide a stronger test for the same hypothesis. Experiment 1A and 1B were preregistered on the Open Science Framework (OSF, http://dx.doi.org/10.17605/OSF.IO/J64SB; Putnam, 2016); the preregistration contains our target sample size, stopping and data exclusion rules, hypotheses, predictions, and analysis plan.

2.1. Method

2.1.1. Subjects

Eighty Amazon Mechanical Turk workers (43 male, 36 female, and 1 other; *M* Age = 34.16; MTurk; www.mturk.com; Buhrmester, Kwang, & Gosling, 2011) received $0.70 for participating in the 15 min experiment. As noted in our preregistration, we planned to omit subjects who reported being non-fluent in English, reported using external resources during the experiment, or who showed a pattern of data that indicated they were not following instructions. However, no subjects met these criteria in Experiment 1A. All subjects were treated in line with the APA ethical guidelines, and the Carleton College IRB approved all of the experiments in this study.

2.1.2. Materials

The materials were a set of 40 trivia claims (20 true and 20 false) adapted from previous research (Fenn et al., 2013; Newman et al., 2012). For each statement we wrote an in-text citation that plausibly supported the claim. For example, "The largest European glacier is Vatnajökull on Iceland (Gudmundsson, 1997)" is a true statement presented with a citation whereas "Baghdad is the capital of Iran" is a false statement presented without a citation. Qualtrics (Provo, UT) was used to present the materials online.

2.1.3. Design and counterbalancing

We used a 2 (citation: yes, no) × 2 (statement truth: true, false) mixed design with *citations* varied between-subjects and *statement truth* varied within-subjects. Half of the subjects saw each statement presented with a citation (40 subjects) and half of the subjects saw each statement presented without a citation (40 subjects).

2.1.4. Procedure

After providing informed consent, subjects began the experiment by answering demographic questions. Then subjects learned that they would see a series of trivia statements presented one at a time, and that their job was to judge the truth of each statement by using a 7-point Likert scale (1 = *Certainly False*, 2 = *Probably False*, 3 = *Possibly False*, 4 = *Uncertain*, 5 = *Possibly True*, 6 = *Probably True*, 7 = *Certainly True*). Subjects were told to respond as quickly as possible, but not so quickly that they made mistakes. After reading the instructions subjects did three practice trials, reviewed the instructions, and then started the experiment.

The trivia statements appeared one at a time, in a random order, in a black font on a white background. The Likert scale appeared as a slider below each statement, with the slider starting on 4 (*Uncertain*). Subjects dragged the slider to their selected number, and clicked a button to submit their response. After all 40 statements had been presented and rated, subjects were asked if they had used any external sources during the experiment. They were then thanked, debriefed, and given a survey completion code to submit on Mechanical Turk for receiving payment.

2.2. Results and discussion

The data for all experiments are available on the Open Science Framework (Putnam, 2016). Fig. 1 displays the average truth ratings for Experiment 1A. A 2 (citation: yes, no) × 2 (statement truth: true, false) mixed-model ANOVA revealed that the true items (M = 4.78, SD = 0.56) received higher truth ratings than the false items (M = 4.02, SD = 0.64), $F(1,78) = 102.42$, $p < .001$, $\eta_p^2 = .57$. *Citations*, however, did not affect the truth ratings, $F(1,78) = 0.50$, $p = .480$, $\eta_p^2 = .01$, indicating that subjects provided similar ratings for items presented with and without a citation. This outcome supported our second hypothesis, that *citations* would not affect believability. To our surprise there was a significant interaction, $F(1,78) = 4.07$, $p = .047$, $\eta_p^2 = .05$, albeit with a small effect size. Fig. 1 suggests that the true items received similar ratings regardless of whether they were presented with a citation or not, but that the false items received lower truth ratings when they were presented with a citation compared to when they were not. This outcome was opposite our prediction that citations might *increase* truth ratings. However, a follow up t-test comparing the false items with citations to the false items without citations was not significant, $t(78) = 1.61$, $p = .110$, $d = 0.36$. One explanation for this unexpected finding is that subjects were relying on preexisting knowledge. If they recognized some of the statements as false, then the presence of a citation might have caused subjects to provide an extremely low rating (e.g. "I know that isn't true, even if there is a citation"). We addressed this concern by using more difficult trivia claims in Experiments 2 and 3.

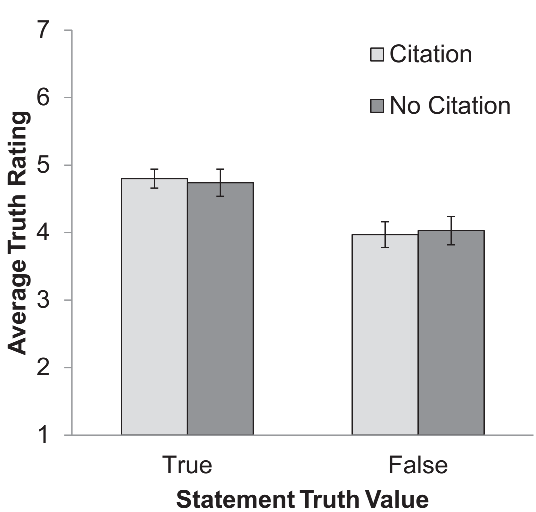

Fig. 1. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation for Experiment 1A. Citation presence was manipulated between-subjects. Error bars depict 95% confidence intervals.

In addition to the above analysis, we also conducted a signal detection analysis to examine bias in responding, which indicates whether there was a tendency to respond true or false (mirroring the analysis done in previous research; Newman et al., 2012).

Following a procedure used by Begg et al. (1992) we converted responses of 5, 6, and 7 on the Likert scale (*Possibly True*, *Probably True*, and *Certainly True*) to have a value of 1, and all responses of 1, 2, 3, and 4 to have a value of 0. Doing so allowed us to calculate the signal detection theory (SDT) measures of c and d' for each subject, which index response bias and response discriminability respectively (Stanislaw & Todorov, 1999). Bias, as measured by c, is independent of d' and in the context of this experiment indicates a tendency to respond "true" or "false." A c value of zero indicates an unbiased response, whereas a negative c value represents a tendency to say true, and a positive c value represents a tendency to say false.[2] The left half of Table 1 shows the SDT measures broken down by whether items were presented with or without a citation (along with the base hit and false alarm rates). An independent samples t-test failed to reject the null, $t(78) = 0.54$, $p = .590$, $d = 0.12$, suggesting that subjects responded similarly for both the items presented with and without a citation (although note that numerically the items with citations showed a stronger truth bias than the items without citations). As an exploratory analysis we also examined whether citations affected d', but again, failed to reject the null hypothesis, $t(78) = 1.15$, $p = .254$, $d = 0.25$, indicating that subjects were equally accurate in their truth judgments for items presented with and without a citation.

Finally, we originally intended to examine whether a subject's education or profession might affect response bias (we noted this as a potential exploratory analysis in our preregistration). We hypothesized that more educated subjects would be more likely to show inflated truth ratings when statements were presented with a citation. However, in Experiment 1A and in all of the remaining experiments there were too few subjects in each category to justify conducting statistical analyses.[3]

3. Experiment 1B

Experiment 1B was run at the same time as Experiment 1A and was identical, except that *citations* was manipulated within-subjects, rather than between; we expected that in-text citations would be more likely to increase truth ratings in a within-subjects design when individuals were exposed to both conditions. Otherwise, the materials and procedure were identical.

[2] Although there are measures such as normalized c (c' which is calculated by dividing c by d'; Macmillan & Creelman, 2004) that do not assume independence between c and d', some of the subjects in this set of experiments showed a d' of zero or a negative d' making c' either impossible to calculate or uninterpretable.

[3] Numerically it did appear that subjects with more advanced degrees were more likely to show a citation effect.

Table 1
Mean hit rate, false alarm rate, discrimination (*d′*), and criterion (*c*) for each condition in Experiment 1A and 1B.

	Experiment 1A		Experiment 1B	
	No citation	Citation	No citation	Citation
Hits	.58 (.17)	.62 (.16)	.62 (.21)	.62 (.22)
False alarms	.45 (.20)	.47 (.17)	.45 (.21)	.47 (.20)
d′	0.35 (.52)	0.48 (.51)	0.52 (.67)	0.46 (.72)
c (bias)	− 0.05 (.50)	− 0.11 (.41)	− 0.11 (.57)	− 0.14 (.53)

Note. Bias represents a tendency to respond "true." Zero represents no bias and negative values indicate subjects were more likely to respond true. Standard deviations are displayed in parentheses.

3.1. Method

3.1.1. Participants

Eighty-two Mechanical Turk workers (47 male, 34 female, and 1 other; *M* Age = 34.85) participated in Experiment 1B (we aimed to recruit 80 subjects, but two subjects likely completed the survey on Qualtrics but failed to submit their payment code on Mechanical Turk). No subjects were excluded from the analyses.

3.1.2. Design and counterbalancing

We used a 2 (citation: yes, no) × 2 (statement truth: true, false) within-subjects design. For each subject half of the statements were presented with a citation (10 true and 10 false) and half were presented without a citation (10 true and 10 false); citation presence was counterbalanced across subjects.

3.2. Results and discussion

Fig. 2 shows the truth ratings for Experiment 1B. Again, the true statements ($M = 4.95$, $SD = 0.57$) received higher truth ratings than the false statements ($M = 4.00$, $SD = 0.61$), $F(1,81) = 176.90$, $p < .001$, $\eta_p^2 = .69$. Replicating Experiment 1A, *citations* did not affect truth ratings, either as a main effect $F(1,81) = 0.04$, $p = .841$, $\eta_p^2 = .001$, or as an interaction, $F(1,81) = 0.58$, $p = .447$, $\eta_p^2 = .01$. Thus, citations did not inflate truth ratings.

The right half of Table 1 displays SDT analyses for the two groups of subjects in Experiment 1B. Critically, there was no difference between conditions in response bias, $t(81) = 0.77$, $p = .447$, $d = 0.05$, again suggesting that citation presence did not affect a tendency to respond true (although again note that the truth bias was larger for the items with citations). Furthermore, there was no difference in *d′*, $t(81) = 0.44$, $p = .659$, $d = 0.09$, indicating that subjects had similar discriminability regardless of whether items were presented with a citation or not. (In future experiments we will report *d′* values, but as they are not the main focus of this paper, will not report any statistical tests.) In sum, both Experiment 1A and Experiment 1B provided support for our secondary hypothesis, that citations would not influence truth ratings.

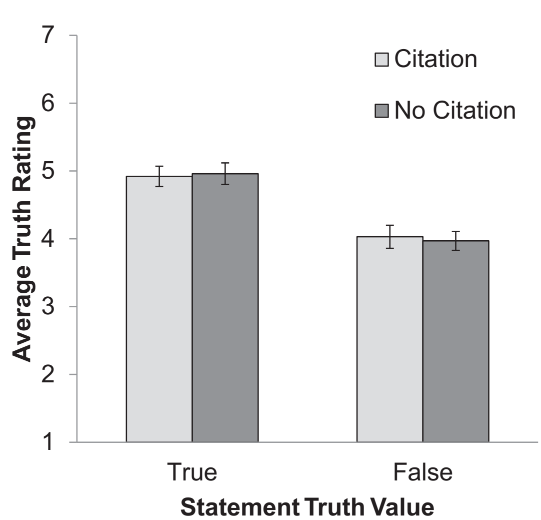

Fig. 2. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation for Experiment 1B. Citation presence was manipulated within-subjects. Error bars depict 95% confidence intervals.

4. Experiment 2A

Experiments 1A and 1B suggested that citations did not increase the believability of trivia claims. One possibility is that subjects might have already known the truth status of some of the trivia claims. Indeed, other research suggests that the illusory-truth effect disappears when the actual truth status of a claim is known (Dechêne et al., 2010). If subjects have prior knowledge about a claim then they will rely on their own knowledge, rather than the presence of an in-text citation to shape their truth judgment. To investigate this hypothesis, Experiments 2A and 2B used a combination of difficult and easy trivia items. We predicted that if citation presence inflated truth ratings, then this effect would be more likely to appear with the hard items, where subjects are less likely to use pre-existing knowledge.

4.1. Method

Experiment 2A was identical to Experiment 1A except for the new materials; otherwise the procedure was identical.

4.1.1. Subjects

We aimed to recruit 80 subjects, but ended up with 82 Mechanical Turk workers. We cut one subject for reporting using external resources, leaving a final sample of 81 (42 male and 39 female; *M* Age = 35.14).

4.1.2. Materials

We created 80 new trivia statements. We selected 40 trivia questions from Tauber, Dunlosky, Rawson, Rhodes, and Sitzman (2013) that had correct recall rates between 0.00 and 0.07 ($M = 0.02$) and 40 questions that had correct recall rates between 0.19 and 0.78 ($M = 0.48$). We converted all of the questions into a statement form (e.g., "Who was the first ruler of the Holy Roman Empire?" became "Charlemagne was the first ruler of the Holy Roman Empire"), and for half of the statements, changed the target word to a plausible, but incorrect response (e.g., "Caesar was the first ruler of the Holy Roman Empire"). Finally, we wrote a plausible citation for each statement. This resulted in a set of 40 hard statements and 40 easy statements, with each set having 20 true statements and 20 false statements.

4.1.3. Design and counterbalancing

Experiment 2A used a 2 (citation: yes, no) × 2 (statement truth: true, false) × 2 (item difficulty: hard, easy) mixed design; *citation* and *item difficulty* were varied between-subjects and *statement truth* was varied within-subjects. Subjects were randomly assigned to one of four groups: one group rated easy statements with citations ($n = 21$), another group rated easy statements without citations ($n = 21$), another group rated hard statements with citations ($n = 19$), and the last group rated hard statements without citations ($n = 20$). Each list that subjects studied was composed of 20 true items and 20 false items.

4.2. Results and discussion

Fig. 3 displays the mean truth ratings; the top panel corresponds to the easy items and the bottom panel corresponds to the hard items. A 2 (citation: yes, no) × 2 (statement truth: true, false) × 2 (item difficulty: hard, easy) mixed-model ANOVA revealed a truth effect, as in Experiment 1; the true items ($M = 5.73$ $SD = 0.99$) received higher

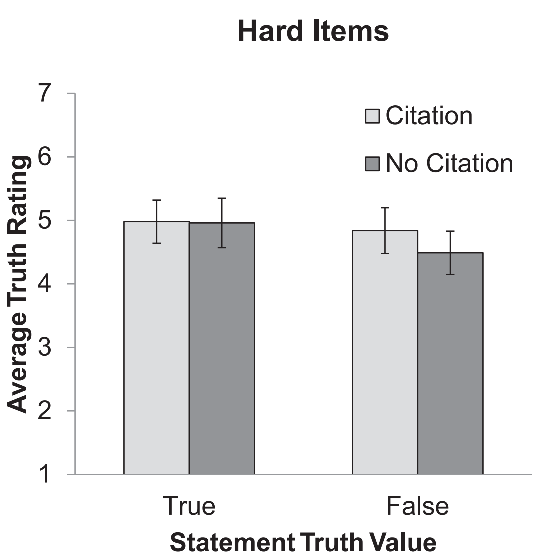

Fig. 3. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation in Experiment 2A. The top panel displays the results for the easy items and the bottom panel displays the results for the hard items. Citation presence was manipulated between subjects. Error bars depict 95% confidence intervals.

truth ratings than the false items ($M = 4.07$, $SD = 1.10$), $F(1,77) = 123.27$, $p < .001$, $\eta_p^2 = .62$. There was also a two-way interaction between *statement truth* and *item difficulty*, $F(1,77) = 80.66$, $p < .001$, $\eta_p^2 = .51$, indicating that subjects had an easier time discriminating the true items from the false items when the statements were easy, compared to when they were hard. For the critical analysis, however, *citations* did not affect truth ratings either as a main effect, $F(1,77) = 0.63$, $p = .431$, $\eta_p^2 = .01$, nor as an interaction with *statement truth*, $F(1,77) = 2.38$, $p = .127$, $\eta_p^2 = .03$. None of the other main effects or interactions were significant (all $Fs \leq 2.26$). Note as well that the odd interaction in Experiment 1A, where the false items with citations received lower truth ratings than the false items without citations, did not replicate.

Table 2 displays the SDT analysis. A 2 (citation: yes, no) × 2 (item difficulty: hard, easy) ANOVA with *c* as the dependent variable revealed a main effect of *item difficulty*, $F(1,77) = 7.58$, $p = .007$, $\eta_p^2 = .090$, suggesting that there was a stronger truth bias for the easy items

Table 2
Mean hit rate, false alarm rate, discrimination (d'), and criterion (c) for easy and hard items as a function of whether items were presented with or without citations in Experiment 2A.

	Easy items		Hard items	
	No citation	Citation	No citation	Citation
Hits	.94 (.05)	.87 (.17)	.58 (.31)	.55 (.27)
False alarms	.38 (.19)	.41 (.23)	.56 (.22)	.65 (.17)
d'	1.89 (.59)	1.54 (.93)	0.07 (1.18)	− 0.28 (.91)
c (bias)	− 0.62 (.67)	− 0.54 (.50)	− 0.21 (.67)	− 0.31 (.56)

Note. Bias represents a tendency to respond "true." Zero represents no bias and negative values indicate subjects were more likely to respond true. Standard deviations are displayed in parentheses.

($M = -0.58$, $SD = 0.44$) than the hard items ($M = -0.26$, $SD = 0.61$). Critically, there was no main effect of *citations* (the means were nearly identical $M_{\text{citation}} = -0.43$, $SD_{\text{citation}} = 0.54$; $M_{\text{no citation}} = -0.42$, $SD_{\text{no citation}} = 0.57$), $F(1,77) = 0.003$, $p = .955$, $\eta_p^2 = .00$, nor an interaction, $F(1,77) = 0.57$, $p = .451$, $\eta_p^2 = .01$. This outcome supported our secondary hypothesis, that citations did not affect truth ratings.

5. Experiment 2B

Experiment 2B was identical to Experiment 2A except that *citations* was manipulated within-subjects, rather than between-subjects. As in Experiment 1, we expected that the within-subjects design would be more likely to show an effect of *citations*.

5.1. Method

5.1.1. Participants

Eighty Mechanical Turk workers participated in Experiment 2B (40 male, 39 female, and 1 other; M Age = 33.89). No subjects were excluded from the analyses.

5.1.2. Design and counterbalancing

Experiment 2B used a 2 (citation: yes, no) × 2 (statement truth: true, false) × 2 (item difficulty: hard, easy) mixed design; *citations* and *statement truth* were varied within-subjects and *item difficulty* was varied between subjects. Subjects were randomly assigned to rate the hard or the easy materials. For each subject half of the statements appeared with a citation (10 true and 10 false) and half appeared without a citation (10 true and 10 false). Each statement appeared with or without a citation equally often across subjects.

5.2. Results and discussion

Fig. 4 displays the mean truth ratings for Experiment 2B with the top panel corresponding to the easy items and the bottom panel corresponding to the hard items. The true items received higher truth ratings ($M = 5.38$, $SD = 0.85$) than the false items ($M = 4.50$, $SD = 0.83$), $F(1,78) = 34.24$, $p < .001$, $\eta_p^2 = .31$. There was also a main effect of *item difficulty*, $F(1,78) = 5.54$, $p = .021$, $\eta_p^2 = .07$, indicating that the easy items ($M = 5.06$, $SD = 0.42$) received higher ratings than the hard items, ($M = 4.82$, $SD = 0.48$). However, both of these main effects were qualified by a significant *statement truth* by *item difficulty* interaction, $F(1,78) = 8.58$, $p = .004$, $\eta_p^2 = .10$, suggesting that subjects were more accurate at rating the easy items than the hard items. Critically, *citations* failed to influence the truth ratings as a statistically significant main effect, although numerically items with citations ($M = 4.99$, $SD = 0.52$) were rated higher than the items without citations ($M = 4.89$, $SD = 0.54$), $F(1,78) = 3.16$, $p = .079$, $\eta_p^2 = .04$. Furthermore, *citations* did not interact with *statement truth* or

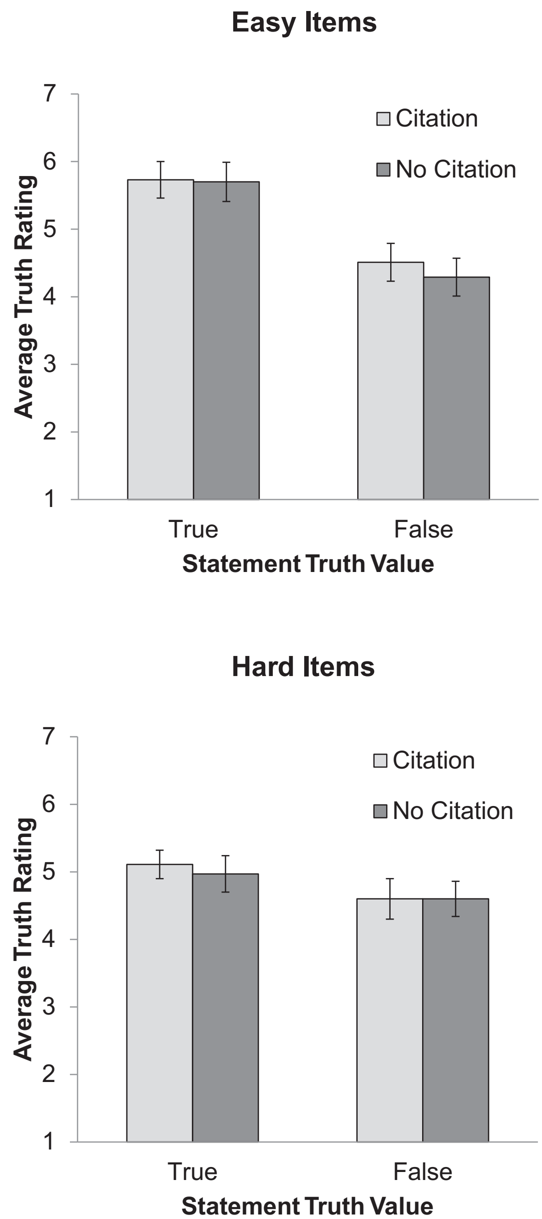

Fig. 4. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation in Experiment 2B. The top panel displays the results for the easy items and the bottom panel displays the results for the hard items. Citation presence was manipulated within-subjects. Error bars depict 95% confidence intervals.

item difficulty (both *F*s < 1). No other main effects or interactions were significant (all *F*s < 3.17).

Table 3 displays the SDT analyses. A 2 (citation: yes, no) × 2 (item difficulty: hard, easy) mixed-model ANOVA with *c* as the dependent variable revealed a main effect of *item difficulty*, $F(1,78) = 6.12$, $p = .016$, $\eta_p^2 = .07$, indicating that subjects showed a larger truth bias when the items were easy than when they were hard.

Critically, we did not find a statistically reliable effect of *citations*, $F(1,78) = 3.81$, $p = .054$, $\eta_p^2 = .05$. However, examining the means and the effect size ($d = 0.21$) suggested that subjects might show a truth bias when items were presented with a citation ($M = -0.52$, $SD = 0.47$) than without a citation ($M = -0.42$, $SD = 0.47$). This was the first indication in all of our experiments that the presence of a citation could inflate truth ratings. The interaction was not significant, ($F < 1$).

Table 3
Mean hit rate, false alarm rate, discrimination (*d′*), and criterion (*c*) for easy and hard items as a function of whether items were presented with or without citations in Experiment 2B.

	Easy items		Hard items	
	No citation	Citation	No citation	Citation
Hits	.77 (.21)	.79 (.18)	.59 (.26)	.67 (.24)
False alarms	.55 (.19)	.59 (.18)	.60 (.20)	.60 (.21)
d′	0.78 (1.20)	0.68 (1.05)	− .01 (1.01)	0.23 (0.98)
c (bias)	− 0.54 (.35)	− 0.63 (.38)	− 0.31 (.54)	− 0.41 (.54)

Note. Bias represents a tendency to respond "true." Zero represents no bias and negative values indicate subjects were more likely to respond true. Standard deviations are displayed in parentheses.

6. Experiment 3A

Four experiments showed that numerically, statements with citations were rated as more likely to be true than statements without citations, but none of these comparisons were statistically significant. The SDT analysis in Experiment 2B, however, hinted that citations might inflate truth ratings in some situations. In Experiment 3A our goal was to conduct a high-powered experiment. First, although power analyses and examining previous research (e.g., Newman et al., 2012) indicated that our initial studies were adequately powered to detect a medium effect size, we increased the sample size to 220 subjects to enable us to detect a smaller effect size. Second, we used the hard item list from Experiment 2, theorizing that if *citations* would have an effect it would be more likely if subjects could not rely on pre-existing knowledge. Third, we manipulated citation presence within-subjects, rather than between, to highlight the contrast between the two statement truths. Finally, we added an instructional manipulation. In the previous experiments subjects were simply told to read the statements and to rate whether they were true or false; nothing was said about the in-text citations. In Experiment 3A a control group received the same instructions as in the previous experiments and an informed group read additional instructions explaining what an in-text citation was, and how authors use them in academic texts. We predicted that informing subjects about the role of citations might increase the weight that citations play in forming judgments of truth.

6.1. Method

Experiment 3A was identical to Experiment 2B except for the addition of the *instructions* manipulation.

6.1.1. Subjects

We aimed to recruit 220 Mechanical Turk workers and ended data collection with 222 subjects. We omitted four subjects who reported using external resources, leaving a final sample size of 218 (100 male and 118 female; *M* Age = 35.91).

6.1.2. Materials, design, and counterbalancing

Experiment 3A used only the hard items from Experiment 2B and had a 2 (citation: yes, no) × 2 (statement truth: true, false), × 2 (instructions: informed, uninformed) mixed design; *citations* and *statement truth* were varied within-subjects and *instructions* was varied between-subjects. Subjects were randomly assigned to the *informed* (108 subjects) or *uninformed* condition (110 subjects). In each condition subjects rated 20 true items and 20 false items, half of which were presented with a citation (the items that appeared with a citation were counterbalanced across subjects).

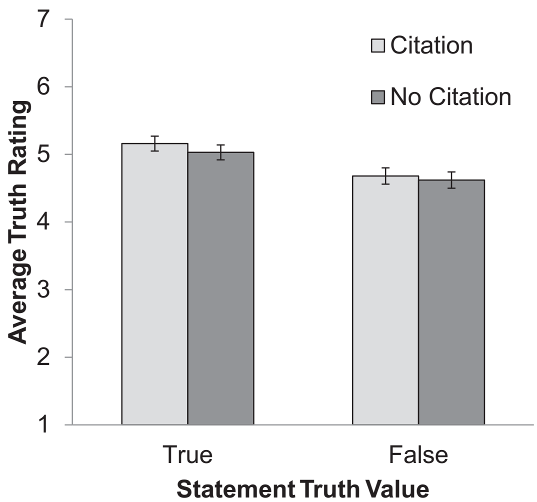

Fig. 5. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation for Experiment 3A. Citation presence was manipulated within-subjects. Error bars depict 95% confidence intervals.

6.1.3. Procedure

The procedure for the uninformed group was identical to Experiment 2B. The procedure for the informed group included an extra paragraph of instructions that informed them that some statements would have a citation and that some would not. Subjects also saw an example of what an in-text citation looked like and read an explanation for how in-text citations are used in academic writing (e.g., "Parenthetical citations are used in academic or scientific writing to refer to the work of others, and are often used to provide supporting evidence for a claim"). After the instructions screen the procedure was identical to Experiment 2B.

6.2. Results and discussion

A 2 (citation: yes, no) × 2 (statement truth: true, false) × 2 (instructions: informed, uninformed) mixed-model ANOVA failed to reveal any main effects or interactions involving *instructions* (all *F*s < 1.06), so Fig. 5 displays the mean truth ratings for Experiment 3A collapsed across *instructions*. We replicated our previous experiments in showing that the true items ($M = 5.09$, $SD = 0.75$) received higher ratings than the false items ($M = 4.65$, $SD = 0.80$), $F(1,216) = 27.99$, $p < .001$, $\eta_p^2 = .115$. Critically, however, we also showed that the items that were presented with a citation ($M = 4.92$, $SD = 0.57$) received higher truth ratings than items presented without a citation ($M = 4.82$, $SD = 0.51$), $F(1,216) = 6.76$, $p = .010$, $\eta_p^2 = .03$. Thus, Experiment 3A demonstrated that the presence of citations increased truth ratings. The *statement truth* by *citations* interaction and the three-way interaction were not significant (both *F*s ≤ 1.06).

The left half of Table 4 displays the results of a SDT analysis, which provided a similar conclusion to the main analysis. A 2 (citation: yes, no) × 2 (instructions: informed, uninformed) mixed-model ANOVA with *c* as the dependent variable showed a main effect of *citations*, $F(1,216) = 8.13$, $p = .005$, $\eta_p^2 = .04$, indicating that subjects had a larger truth bias for the items presented with a citation ($M = -0.38$, $SD = 0.54$) than the items presented without a citation ($M = -0.29$, $SD = 0.47$). Thus, citations did influence truth judgments. There was no main effect of *instructions*, $F(1,216) = 1.65$, $p = .201$, $\eta_p^2 = .008$, nor an interaction, $F(1,216) = 0.52$, $p = .472$, $\eta_p^2 = .002$.

In sum, Experiment 3A suggested that in-text citations increased truth ratings. Why did this pattern occur here but not in the previous experiments? One possibility is that the true effect size is small ($d = 0.18$), and that the larger sample size of Experiment 3A was necessary to detect the effect. Another possibility is that the effect only emerges when certain conditions are met (difficult materials and manipulating citations within subjects). Surprisingly, both the informed and uninformed instructional groups yielded similar truth ratings. This outcome suggests that subjects either already have a good sense of what in-text citations are used for, or that the instructions did not change how subjects were weighting the presence of citations when making their truth judgments.

Table 4
Mean hit rate, false alarm rate, discrimination (*d*′), and criterion (*c*) for each condition in Experiment 3.

	Experiment 3A		Experiment 3B	
	No citation	Citation	No citation	Citation
Hits	.59 (.23)	.63 (.25)	.61 (.23)	.63 (.24)
False alarms	.60 (.20)	.61 (.21)	.57 (.22)	.59 (.22)
d′	0.02 (0.98)	0.11 (1.04)	0.13 (.94)	0.18 (.89)
c (bias)	− 0.29 (.47)	− 0.38 (.54)	− 0.31 (.58)	− 0.37 (.63)

Note. Bias represents a tendency to respond "true." Zero represents no bias and negative values indicate subjects were more likely to respond true. Standard deviations are displayed in parentheses.

7. Experiment 3B

Experiment 3B was a direct replication of Experiment 3A.

7.1. Method

The method was almost identical to Experiment 3A: The only change to the procedure was that at the end of the experiment we asked subjects to describe how in-text citations are used in academic writing in an open-ended response field. We aimed to recruit 220 subjects from Mechanical Turk and ended up with data from 221. One subject was excluded from all analyses for reporting using external resources, leaving a final sample size of 220 (118 male and 102 female; *M* age = 37.1). There were 111 subjects in the uninformed condition and 109 in the informed condition.

7.2. Results and discussion

We ran a 2 (citation: yes, no) × 2 (statement truth: true, false) × 2 (instructions: informed, uninformed) mixed-model ANOVA. As seen in Fig. 6, the true items ($M = 5.13$, $SD = 0.75$) received higher ratings than the false items ($M = 4.61$, $SD = 0.76$), $F(1,218) = 43.72$, $p < .001$, $\eta_p^2 = .167$. Critically, there was not a statistically significant effect of *citations*, although numerically trivia statements with citations ($M = 4.91$, $SD = 0.61$) were rated higher than trivia statements without citations ($M = 4.84$, $SD = 0.53$), $F(1,218) = 3.12$, $p = .079$, $\eta_p^2 = .014$. As in Experiment 3A there was no main effect of *instructions*, ($F < 1$), however in contrast to Experiment 3A there was a significant interaction between *instructions* and *citations*, $F(1,218) = 8.29$, $p = .004$, $\eta_p^2 = .037$. Subjects in the uninformed group provided similar ratings for both items presented with a citation ($M = 4.85$, $SD = 0.53$) and without a citation ($M = 4.90$, $SD = 0.52$), $t(110) = 0.94$, $p = .352$, $d = 0.06$, whereas subjects in the informed group provided higher ratings for the items presented with a citation ($M = 4.97$, $SD = 0.68$) than without a citation ($M = 4.78$, $SD = 0.54$), $t(108) = 2.88$, $p = .005$, $d = 0.31$. Thus, it appears that in Experiment 3B the instructions caused subjects to weight citations more heavily when making their judgments of truth.

The three-way interaction and the other two-way interactions were not significant, (all *F*s ≤ 1.73).

The right half of Table 4 displays the results of a SDT analysis, which provided a similar conclusion to the main analysis. A 2 (citation: yes, no) × 2 (instructions: informed, uninformed) mixed-model ANOVA with *c* as the dependent variable did not show a main effect of

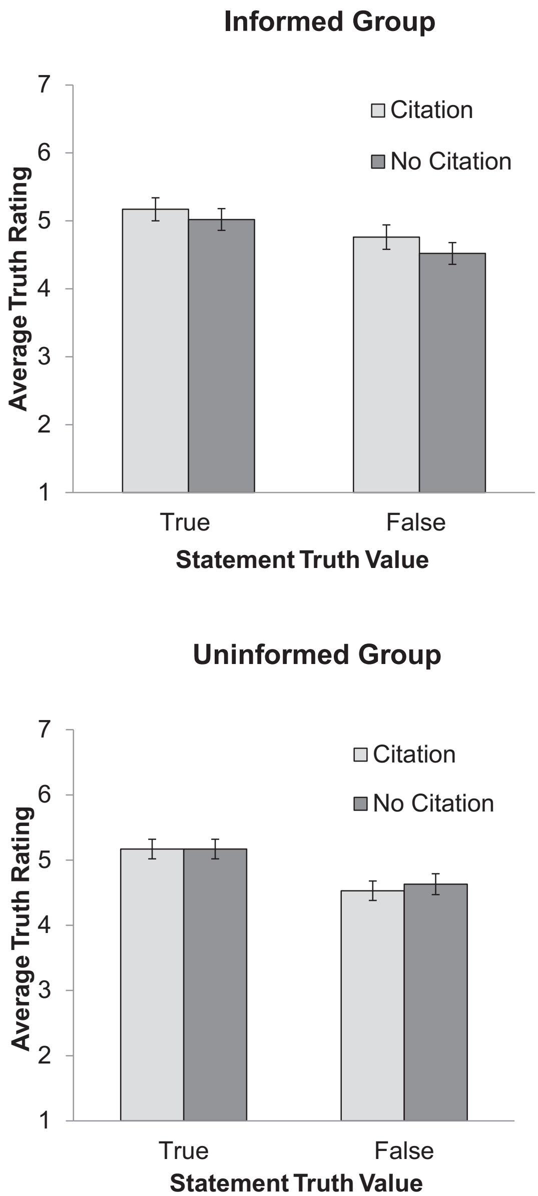

Fig. 6. Average truth ratings as a function of whether the claims were true or false and whether they were presented with or without a citation for Experiment 3B. The citation presence was manipulated within-subjects. Subjects in the uninformed group provided similar ratings for both items with and without a citation whereas subjects in the informed group provided higher truth ratings for items with a citation compared to items without a citation. The error bars depict 95% confidence intervals.

citations, $F(1,218) = 2.90$, $p = .090$, $\eta_p^2 = .01$, suggesting that subjects had a similar bias to respond true for the items presented with a citation ($M = -0.37$, $SD = 0.63$) to the items presented without a citation ($M = -0.31$, $SD = 0.58$); this was a different outcome from Experiment 3A, although numerically was in the same direction. There was no main effect of *instructions*, $F(1,218) = 1.35$, $p = .246$, $\eta_p^2 = .006$. However, there was a significant interaction, $F(1,218) = 7.12$, $p = .008$, $\eta_p^2 = .03$, (again this finding was inconsistent with Experiment 3A). Follow-up analyses suggested that the subjects in the uninformed group had a similar truth bias for both the items with citations ($M = -0.36$, $SD = 0.59$) and without citations ($M = -0.39$, $SD = 0.59$), $t(110) = 0.76$, $p = .449$, $d = 0.05$, but that subjects in the informed group had a larger truth bias for the items presented with a citation ($M = -0.37$, $SD = 0.67$) than without ($M = -0.21$, $SD = 0.55$), $t(108) = 2.82$, $p = .006$, $d = 0.26$. Given that this outcome only occurred in Experiment 3B, it must be interpreted with caution. However, the interaction appears to be driven by a reduction in the truth bias for the items presented without a citation, rather than an increase in the truth bias for the items presented with a citation suggesting that informed subjects were more skeptical of items presented without a citation.

8. Meta-analysis of all experiments

Given the similar design and mixed results of the experiments reported here, we conducted a mini meta-analysis to provide a quantitative estimate of the effects of citation presence on truth ratings (readers who are interested in the details of conducting a mini meta-analysis should consult Cummings, 2012). Cummings (2012) argues that meta-analysis, even when conducted on a small scale as in the current study, can lead to large increases in precision of measuring effects and can help clarify ambiguous results. We used bias, or *c*, as our main dependent variable, and calculated an effect size (Cohen's *d*) and 95% confidence interval for each experiment, coded such that a larger effect size represented a larger truth bias for statements presented with a citation than without a citation.[4] Thus, an effect size of 0 would indicate a similar truth bias for items presented with a citation and without a citation. We then calculated weighted effect sizes for each experiment. When calculating the weighted effect size for the within-subjects experiments we took into account the correlation between the bias scores for the items with and without citations to account for the fact that the same subject was providing those two different ratings (Borenstein, Hedges, Higgins, & Rothstein, 2009). We used a random effects model for the final analysis (and used the metafor package in R to run the analysis (Viechtbauer, 2010)).[5] Fig. 7 displays a forest plot with the effect size for each experiment and the results of the meta-analysis at the bottom. The model estimated that the overall effect size for the effects of *citations* on perceived truthfulness was $d = 0.13$, 95% CI [0.06, 0.20], indicating that there is a very small, but consistent effect of *citations*; subjects tended to show a larger truth bias for the statements presented with a citation than without a citation.

9. General discussion

The goal of these experiments was to determine how much in-text citations affect the perceived "truthiness" of trivia claims. Despite some variability in results across the experiments, the mini meta-analysis (see Fig. 7) supports the hypothesis that there is a small *citation effect*: subjects were more likely to rate a statement as true if it was presented with a citation than without a citation. This is the first time, to our knowledge, that citations have been shown to lead to higher judgments of truth, albeit with a small effect size ($d = 0.13$).

At a surface level, the current results parallel the finding that nonprobative photos inflate truthiness (e.g., Fenn et al., 2013; Newman et al., 2012). Despite the surface level similarity, in-text citations are likely affecting truth judgments via a different mechanism. Nonprobative photos–by definition–do not provide helpful information about a claim, whereas citations ostensibly are probative–they point the reader to an external source that should provide evidence for a claim (although note that a critical reader will investigate a particular reference before believing an author's claim, e.g., is it a peer-reviewed journal article or a personal blog post?). Furthermore, photos are vivid and visually appealing, whereas citations are not.

Nonprobative photos are thought to inflate truth ratings by increasing both the type and amount of information that people retrieve

[4] We conducted a second meta-analysis where the effect size was calculated based on the truth ratings provided by subjects (rather than *c*). The results of this meta-analysis led to similar conclusions.

[5] Although we used a random effects model for the meta-analysis, a fixed effect model would have been appropriate; our *Q* value was 2.54 and I^2 was zero, a strong indicator that there was homogeneity among the estimates.

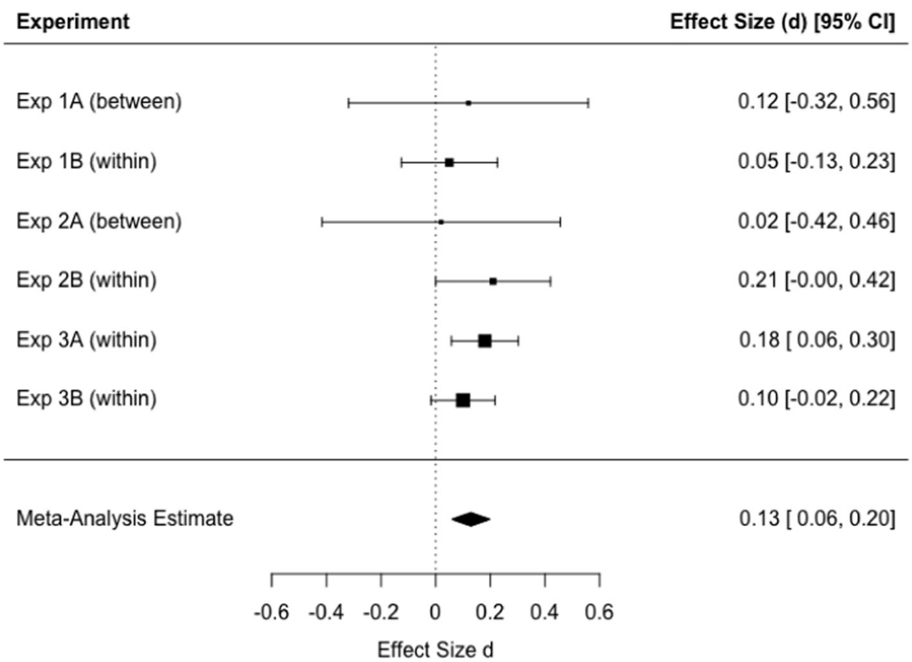

Fig. 7. Forest plot depicting effect sizes and 95% confidence intervals around the estimated effect sizes. A positive effect size (Cohen's *d*) indicates a larger truth bias (as measured by *c*) for statements presented with citations than without citations.

and by increasing the fluency with which a trivia claim is processed. Seeing a photo might cause people to retrieve more related information to the target statement, or people might selectively search the photo to confirm a hypothesis about whether a claim is true or not (Fenn et al., 2013; Newman et al., 2012, 2015). Cardwell et al. (2016) suggested that processing photos is a relatively fluent experience (compared to text) and that the fluency experience is misattributed to the statement being true; this is similar to what happens in the truth effect where the fluency and familiarity of seeing an item a second time also inflates truth ratings (Begg et al., 1992; Hasher et al., 1977; see Alter & Oppenheimer, 2009 for a review on fluency).

In contrast to photos, citations are not visually appealing, and if anything might *decrease* fluency rather than increase it. Reading an article introduction that is packed with in-text citations parallels the cognitive control task where subjects read a story with distracting words inserted in a different type font (e.g., *The car ride* river *was getting bumpy* jeep *now*…; Connelly, Hasher, & Zacks, 1991). Furthermore, in contrast to nonprobative photos, citations are a legitimate piece of evidence to consider when evaluating the truth of a statement. Thus, it seems likely that citations are not inflating truth judgments via fluency, but that instead, readers are likely using citations as a form of evidence.

In sum, citations appear to have a small effect on truth judgments. Given that citations are probative, a natural question is why the effect is not larger. One possible explanation is that the subjects in our experiments did not know what in-text citations are used for. However, this does not appear to be the case. At the end of Experiment 3B we asked subjects to briefly describe the purpose of in-text citations, and 90% of their responses included something about referencing external sources or providing support for a claim. Critically, this pattern of responding was identical for the informed and uninformed groups, suggesting that subjects understood what citations are used for without explicit instructions.

A second more plausible explanation is that even if subjects understand what citations are used for, the presence of the citation may not be weighed heavily in their truth judgment. Metacognition researchers have suggested that people draw on a variety of cues when making judgments, including a priori theories or beliefs about how cognitive processes work and their subjective experience of engaging in a cognitive process (Koriat, 1997). However, Koriat, Bjork, Sheffer, and Bar (2004) have shown that subjects typically weigh subjective experiences (such as the fluency experienced when reading an item) much more heavily than general theories about cognitive processes unless they are explicitly encouraged to use theoretically driven cues. Thus, even if subjects know that citations are probative, subjects may not use that information when making their judgments. Furthermore, if citations decrease the fluency of reading a claim, then that disfluency is working against the probative value of the citation, which would lower the size of the citation effect. Future research could explore how controlled and automatic processes contribute to truth judgments (similar to Begg et al., 1992) and begin to examine how subjects weigh different cues when making judgments.

Finally, a third explanation for why the citation effect is not stronger in the current experiments is that some subjects may simply have decided to ignore the citations while reading. As the citations were paired with both true and false statements, subjects may have deduced that citations were not diagnostic of truth, and thus chose to ignore the citations as they read. This would lower the size of the citation effect, but also suggests that our obtained effect size is an underestimate of what the citation effect might look like in the real world.

One unexpected finding in our experiments was the inconsistent effect of instructing people about the importance of citations: in Experiment 3A reminding subjects about the purpose of citations did not influence truth judgments, but in Experiment 3B, subjects who were informed about the role of citations showed a citation effect. Because this pattern did not replicate across the two experiments it is difficult to draw many firm conclusions. One would expect the instructions to lead to a larger citation effect: either the instructions are reminding subjects to weight the presence of a citation more heavily when making a truth judgment (e.g., Koriat et al., 2004), or the instructions could create a demand effect where subjects lower their truth judgments for statements presented without citations. In any case, future research is needed to better understand how reminding people about the role of citations affects how people make truth judgments.

Understanding how and why citations influence perceived truthfulness is important for several reasons. From a cognitive perspective, in-text citations provide another example of how truth ratings can be biased. Citations are ostensibly–but not necessarily–probative, which might offer new directions for research (e.g., can you train people to question the quality of the citation?). Future research could also explore whether citations have a similar effect on perceptions of truth when embedded in a longer text passage or how the fluency of names in citations influence truth judgments. Other research, for example, indicates that people are more likely to believe a claim presented by someone with an easy to pronounce name than a difficult to pronounce name (Newman et al., 2014), leading to the interesting question of whether it is more persuasive to cite an author with an easy to pronounce last name than a difficult to pronounce one.

From an educational perspective, it is important to know how readers respond to the use of citations. Is a scientific article filled with in-text citations more persuasive or believable than one that contains only a few select references? Does the quality or name recognition of the references affect believability? Authors use in-text citations for many reasons, sometimes to support a claim and sometimes just to point interested readers to further resources. Regardless, in-text citations may increase believability, which can have advantages in some situations and disadvantages in others. The experiments reported here are a first step in understanding how and why citations might affect believability. Furthering that understanding will help authors to become more persuasive and readers to become more critical consumers of text.

Acknowledgments

We thank K. Andrew DeSoto and Julia Strand for providing feedback on this manuscript, Jae Eun Lee, and Lucia Ray for their assistance, and Carleton College for supporting this research.

Appendix A. Supplementary materials

Supplementary materials to this article, including materials, data, and analysis scripts are available at the Open Science Framework, osf.io/j64sb, and at http://dx.doi.org/10.1016/j.actpsy.2017.07.010.

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.

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*, 446–458.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical murk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*, 3–5.

Cardwell, B. A., Henkel, L. A., Garry, M., Newman, E. J., & Foster, J. L. (2016). Nonprobative photos rapidly lead people to believe claims about their own (and other people's) pasts. *Memory and Cognition*, *44*, 883–896. http://dx.doi.org/10.3758/s13421-016-0603-1.

Colgrove, J., & Bayer, R. (2005). Could it happen here? Vaccine risk controversies and the specter of derailment. *Health Affairs*, *24*, 729–739.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, *6*, 533–541. http://dx.doi.org/10.1037/0882-7974.6.4.533.

Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

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. http://dx.doi.org/10.1177/1088868309352251.

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

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*, 107–119. http://dx.doi.org/10.1037/0003-066X.47.5.670.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality and Social Psychology*, *59*, 601–613.

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

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656. http://dx.doi.org/10.1037/0096-3445.133.4.643.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, *107*, 343–352.

Michael, R. B., Newman, E. J., Vuorre, M., Cumming, G., & Garry, M. (2013). On the (non)persuasive power of a brain image. *Psychonomic Bulletin and Review*, *20*, 720–725. http://dx.doi.org/10.3758/s13423-013-0391-6.

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.

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

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

Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster, J. L., Bernstein, D. M., & Garry, M. (2014). People with easier to pronounce names promote truthiness of claims. *PloS One*, *9*, e88671–5. http://dx.doi.org/10.1371/journal.pone.0088671.

Putnam, A. L. (2016, July 7). *How do academic citations affect the perceived "truthiness" of statements?* (Preregistration Retrieved from DOI 10.17605/OSF.IO/DV8FS).

Putnam, A. L., Sungkhasettee, V., & Roediger, H. L. (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science*, *11*, 652–660.

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

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of Philosophy and Psychology*, *1*, 563–581.

Scherer, M. (2017, April). Can President Trump handle the truth? *Time*, *189*(12) (doi:http://time.com/4710456/donald-trump-time-interview-truth-falsehood/).

Schwarz, N. (2015). Metacognition. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology, volume 1: Attitudes and social cognition* (pp. 203–229). Washington, DC: APA. http://dx.doi.org/10.1037/14341-006.

Sparks, J. R., & Rapp, D. N. (2011). Readers' reliance on source credibility in the service of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 230–247. http://dx.doi.org/10.1037/a0021331.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149. http://dx.doi.org/10.3758/BF03207704.

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, *31*, 176–203. http://dx.doi.org/10.1080/07370008.2013.769994.

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. http://dx.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.

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

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*, 1–48 (Retrieved from: http://www.jstatsoft.org/v36/i03/).