The tweeter matters: Factors that affect false memory from Twitter

Nicholas R. Griffin*, 1, Cory R. Fleck 2, Mitchell G. Uitvlugt, Susan M. Ravizza, Kimberly M. Fenn

Department of Psychology, Michigan State University, 316 Physics Rd, East Lansing, MI, 48824, USA

Abstract
Social media websites have dramatically increased in popularity. Information on these sites does not require verification and may be inaccurate. We investigated how familiarity with content and Twitter authors influenced false memory. Participants from Michigan State University (MSU) read an article about college basketball followed by a Twitter recap that contained false information. The article discussed either MSU or the University of Iowa; the Twitter feed was purportedly written by students from one of the schools in a 2 × 2 design. Lastly, participants completed recognition and source memory tests. Correct and false recognition were similar across conditions. In an analysis including source test results, we found higher false memory for the Iowa article than the MSU article. We also found an interaction between article and Twitter author, opposite of our hypothesis, such that false memory was greatest for conditions with incongruent article and Twitter author (e.g., Iowa article and MSU Twitter). Thus, we replicated earlier work showing higher memory distortion for unfamiliar content. We also found preliminary evidence that readers may take into account the relationship between authors and content. This is an important first step in understanding how false memory arises from social media, an everexpanding cultural influence.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction
Social media is becoming an increasingly popular source of information. Originally, social media sites were a place for sharing personal information (e.g., life events, anecdotes); however, a recent report indicated that 62% of Americans now read news about current events from social media outlets (Gottfried & Shearer, 2016). Using social media for disseminating news has unique advantages. For example, smartphones allow for events to be transmitted to social media platforms like Twitter in real time with continuous updates. Interestingly, approximately 24% of U.S. Internet users regularly visit Twitter, and 59% of those individuals use Twitter to obtain news (Duggan, Page, & Manager, 2016). Furthermore, of the top five most used social networks, Twitter has the greatest percentage of users who use the platform for the express purpose of reading news, rather than happening to read news while completing other tasks (Gottfried & Shearer, 2016).

Although accurate and direct accounts of an event can be posted to Twitter in real time, individuals can also post information without any direct experience or knowledge of a given event. That is, unlike traditional news sources, individuals may post information that they have not actually witnessed without any oversight to assure its validity. The lack of oversight on Twitter means that, potentially unbeknownst to other users, information may not be factually accurate. Although false information may be corrected in subsequent posts (Mendoza, Poblete, & Castillo, 2010), this false information remains accessible, which may have negative consequences for memories of events (Jin et al., 2014; Starbird, Maddock, Orand, Achterman, & Mason, 2014). In the present study, we were interested in false information on Twitter and factors that influence the extent to which erroneous information is integrated into memory.

Decades of research have shown that memory is not a pure account of experienced events and is susceptible to distortion through inaccurate or misleading information that is subsequently presented (Loftus, Miller, & Burns, 1978). This contradictory...
information (i.e., misinformation) may be integrated with or over-ride original memories, resulting in distortion (Laney & Loftus, 2013; Loftus, 2005; Schacter, Guerin, & St. Jacques, 2011). Recent evidence suggests that false information garnered from another individual can also lead to false memories (Gabbert, Memon, Allan, & Wright, 2004; Roediger, Meade, & Bergman, 2001). A related study found that if false information was communicated orally from another person, false memory in subsequent testing was greater than if the exact same information was read from a media report or leading questions, or if participants received no false information (Paterson & Kemp, 2006). However, there appears to be some nuances to the interpersonal transmission of false information from the latter study, as evinced through an additional experimental group in which participants read false information that they believed originated from a co-witness of the same event. The false memory results from this written co-witness condition, although numerically greater than from the media report and leading question conditions, were not significantly different. Therefore, it appears that socially transmitted false information may be more likely to be integrated into memory than misinformation presented through different media.

Social media may provide a unique platform for the interpersonal transmission of false information. Given that memory distortion may increase when misinformation is presented by another individual, we investigated the effect of information dissemination through social media on subsequent memory distortion. Our previous work found that individuals are less likely to incorporate false information from Twitter than from a control condition that resembled Twitter (Fenn, Griffin, Utvglut, & Ravizza, 2014). In that study, participants viewed images that depicted a story (e.g., Okado & Stark, 2005) and read statements recapping those images. Some statements contained misinformation about the original events. The critical manipulation was the medium used to present misinformation. Participants viewed the statements in either a Twitter feed purportedly written by previous participants or a less familiar format that resembled the Twitter feed but was not associated with social media. Participants rated their confidence that correct and false items were present in the images on an eight-point scale from 1 “Definitely did not see in pictures” to 8 “Definitely saw in pictures.” Interestingly, participants in the Twitter condition had significantly lower ratings for false items than controls, suggesting that their mean was closer to the lower end (did not see in pictures) than the control group. Our primary finding may be due to increased skepticism or lower perceived credibility of Twitter compared to the control feed. However, because both sources purportedly compiled responses from multiple authors, another variable that could have affected the results is inferred differences between individual authors and their relationship to the information. For example, participants may have attributed the misinformation in the recap statements to a subset of previous participants who did not pay close attention to the images. We chose to focus on the relationship between authors and information to allow for closer investigation of the increased skepticism of false information from Twitter.

The relationship individuals have with information may impact the extent to which false information is integrated into memory. For instance, one study asked Yale and Princeton students to indicate agreement with false statements within identical fictional stories set at either Princeton or Yale (Prentice, Gerrig, & Bailis, 1997). The stories contained real-world, common-knowledge statements, some of which were true (e.g., “Sunlight is bad for your skin”), and some of which were false (e.g., “Most forms of mental illness are contagious”). Participants reported greater agreement with false statements that were embedded in a story about the other school than a story about the school they attended. The authors suggest that the relatively high familiarity of the same-school information may underlie the response differences; participants may have more easily rejected false statements due to the easily identifiable conflict with their everyday environment. In contrast, reduced familiarity of the other school environment may have decreased item rejection because of reduced certainty of the falsehood of the statements. Therefore, the origin of a given piece of information may play a critical role in subsequent endorsement of that information as either true or false.

In the present study, we were interested in using Twitter to investigate the relationship between familiarity and false memory to understand how false information from Twitter may lead to memory distortion. We manipulated both the content of the information and the identity of the individuals who provided the false information, via Twitter. We first hypothesized a main effect of content, such that false memory would be higher when participants read a story about a less familiar university (see Prentice et al., 1997). If our previous finding of low false memory for tweets was due to participants’ beliefs about author credibility (see Fenn et al., 2014), then false memory should be higher when Twitter authors had high familiarity with the content. That is, we hypothesized that false memory would be greater when a Twitter source (i.e., the purported author of the tweets) was more familiar with the information (i.e., students from a given university writing about events at their university) than when the Twitter source was less familiar with the information (i.e., students from a given university writing about events at a different university).

2. Methods

2.1. Sample

We recruited 156 undergraduate psychology students (112 females; mean age = 19.5, SD = 1.5) from Michigan State University. Participants received credit in their psychology course for participating.

2.2. Design

We investigated the relationship between false memory and familiarity, with respect to Twitter, in two ways. First, we manipulated familiarity with the information presented at encoding. Second, we manipulated familiarity between social media source (e.g., individuals who post to Twitter) and the information they discussed on Twitter. To do this, we used a modified version of the misinformation paradigm (Loftus et al., 1978). Participants completed three experimental phases (encoding, misinformation, and test), but this study was unique in two ways: it compared information encoded from multiple text sources and presented misinformation through a simulation of Twitter.

To manipulate familiarity with content, participants read a news article either about their home university — Michigan State University (MSU, high familiarity) — or an unaffiliated university — the University of Iowa (Iowa, low familiarity). They then read a Twitter feed about the article which included misinformation. The purported authors of the tweets were manipulated such that they were either familiar with the content (e.g., MSU students tweeting about MSU events) or unfamiliar (e.g., MSU students tweeting about Iowa

---

3 Two participants completed the experiment but were excluded from all analyses because they did not follow experiment instructions. Two additional participants were excluded from the analyses because of their extreme outlier false alarm rates for our significant measure of false memory (false alarms to suggested items in both the recognition and source tests; extreme outlier = IQR + 3 IQR).
events and vice versa). We used a counterbalanced 2 (MSU or Iowa article) × 2 (tweets by MSU or Iowa students) design. Lastly, participants were tested on their memory for the original article and the source where they acquired the information presented during the memory test.

3. Procedure

3.1. Preliminary task

This task established credibility of our “student-composed” Twitter feed that participants would later read. Upon arrival, participants were told that the experiment was investigating Twitter and memory and that they would read two news articles and compose tweets about each article. The articles were based on real press releases posted on university websites: one from MSU about electric car technology, the other from Iowa about flight safety. The articles were edited to be approximately the same length.

Before reading, participants were told that their tweets needed to be identifiable as originating from an MSU student for future research purposes. Participants therefore created their own Twitter profile, selecting one of four MSU-themed icons (e.g., Spartan helmet, MSU seal) and creating an MSU-related username (or Twitter “handle”; e.g., @sparty27). They then viewed each article individually and wrote three tweets in response to each (e.g., “Thoughts on battery operated cars? #electricity #batteries”; “Making flying safer ... the UI way #safety”). Participants were not tested on their memory in this task, and article order was counterbalanced across individuals within each condition.

3.2. Encoding

Participants then read a third news article that discussed a college basketball coach being recruited by an NBA team. Two versions of the article were created — Each featured the respective school’s basketball coach and appeared to be written by that university’s newspaper. One version featured Tom Izzo, the MSU coach; the other featured Fran McCaffery from Iowa (see Fig. 1 in the Supplementary Online Material [SOM]). The story was essentially the same in each version, stating that the Boston Celtics needed a new coach, they recruited Izzo or McCaffery (depending on condition), who declined, and another coach was selected. Only minor details pertaining to the coach’s employment at their respective university (e.g., duration of employment, tournament records) differed between articles, keeping each article mostly factual. The premise that the Boston Celtics recruited a new coach shortly before we began data collection was true; however, the recruitment of these specific coaches was not. To determine whether the participants (all from MSU) showed increased skepticism about the MSU article, participants rated their trust of the information in this article.

Participants were given as much time as they needed to read the article and told to read carefully before continuing.

3.3. Distractor task

After encoding, participants completed the Operation Span (OSPA) task (Unsworth, Heitz, Schrock, & Engle, 2005) to reduce rehearsal. Here, participants judged the veracity of simple arithmetic problems and viewed letters presented individually on the screen. After a set of problems, participants were asked to recall the letters in the order they were presented.

3.4. Misinformation feed

Following the OSPAN, participants read tweets that were purportedly written by prior participants either from MSU or Iowa, depending on condition. The researcher then loaded a website that appeared identical to the research lab website and started the Twitter feed, which was displayed for 3 min and 20 s.

The Twitter feeds contained 36 lines of text that summarized the news article (see Appendix A in SOM). The tweets were written with informal syntax, and some lines used “hashtags” (#) or “at signs” (@) which are common on Twitter. Most information in the feeds was consistent with the article; however, six lines directly conflicted with the original information. Participants were not forewarned of conflicting information.

Two feeds were created for each version of the coach recruitment story: one purportedly authored by MSU students and another by Iowa students. Altogether, there were four versions of the Twitter feed. We told participants that the feeds were written by prior participants, and we designed the feeds to bolster this claim. The feeds purportedly written by MSU students included the four icons from the preliminary task and all Twitter handles were MSU-related (see Fig. 2). Similarly, the Iowa feeds included four relevant icons (e.g., University of Iowa seal, Iowa Hawkeyes logo) and Twitter handles that related to Iowa (e.g., “@BlackAndGold1847”).

Each feed used a background design visually similar to Twitter and displayed four lines of text (i.e., four tweets) at a time. The tweets first appeared in the top row on the page and scrolled down to each lower row after being presented for 5 s. Once a tweet was presented on the bottom row for 5 s, it faded to white. The process continued until all 36 tweets were presented. Most secondary features from Twitter were retained for aesthetic accuracy, including a faux settings menu, buttons labeled “Expand” at the bottom of each tweet, and the Twitter logo at the top of the window (Fig. 2).

3.5. Test

Directly after the misinformation feed, participants rated their attention to the article and Twitter feed, each on a 5-point Likert scale. Participants then completed a recognition memory test, followed by a source test. During the recognition test, 36 items were randomly presented that may or may not have appeared in the main article (see Appendix B in the SOM). The test contained 20

Fig. 1. Proportion of false recognition (±1 SE) by article and Twitter sources for suggested items reported as present in the article during recognition and source tests.
correct items—10 items that appeared only in the article and 10 items that appeared in both the article and Twitter feed. The test also contained 16 items of false information—six items that appeared in the Twitter feed (suggested information) and 10 items that never appeared (novel lures). Each item was presented individually, and participants were asked to determine whether each item “appeared in the article” or “did not appear in the article.” No time limit was imposed. For items that appeared in the article only, or in the article and Twitter feed, a response of “appeared in the article” was considered a hit. If participants responded “did not appear in the article” to these items, it was considered a miss. In contrast, a response of “did not appear” to suggested items and novel lures was considered a correct rejection and a response of “appeared in the article” was considered a false alarm. Immediately following each response, participants rated their confidence on a 5-point Likert scale, ranging from “not at all confident” to “extremely confident.” After the recognition test, participants rated their trust in the information from the article and Twitter, each on a 5-point Likert scale.

Finally, participants completed a source test. Each item from the recognition test (except novel lures) was presented again, and participants were asked to identify where they acquired the information. Five options were presented: “I saw it in the article only,” “I saw it in the Twitter feed only,” “I saw it in both [the article and Twitter feed] and they were the same,” “I saw it in both and they were different,” and “I guessed.”

4. Results

All analyses used a 2 \times 2 ANOVA with article content (MSU-Izzo; Iowa-McCaffery) and Twitter source (MSU tweeters; Iowa tweeters) as between-subjects factors, and used partial eta squared to assess effect size. We first analyzed correct recognition (hits) for items that appeared in both the original article and Twitter feed and found that performance was similar across conditions. We found no main effect of article, \( F(1, 152) = 0.01, p = 0.97 \), or Twitter source, \( F(1, 152) = 0.19, p = 0.66 \), and there was not a significant interaction between the factors, \( F(1, 152) = 0.34, p = 0.56 \). We next analyzed items present only in the article, and we found a small but non-significant effect for higher correct recognition for participants who read the MSU article (71.3%) than the Iowa article (64.7%), \( F(1, 152) = 2.97, p = 0.087, \eta_p^2 = 0.02 \). There was neither a significant effect of Twitter source, \( F(1, 152) = 0.01, p = 0.92 \), nor a significant interaction, \( F(1, 152) = 0.13, p = 0.72 \).

We next tested false alarms to suggested items and novel lures. For suggested items, we found no main effects of article, \( F(1, 152) = 0.34, p = 0.56 \), or Twitter source, \( F(1, 152) = 0.00, p = 0.99 \), and no significant interaction, \( F(1, 152) = 0.03, p = 0.87 \). Similarly, for novel lures, there were no main effects of article or Twitter source, and there was not an interaction, all \( F's < 0.67 \) and \( p's > 0.40 \). We also analyzed confidence ratings (detailed in SOM), and those results did not inform the main results.

Finally, we explored a stronger measure of false memory—the percentage of suggested items that participants reporting seeing in the article. In this measure of false memory, participants had to initially false alarm (in the recognition test) to the item and then, in the subsequent source test, participants had to report seeing the information in the original article rather than correctly reporting that the item was present only in the Twitter feed. This measure is a strong test of the integration of false information with the original memory because it counts suggested items which were falsely endorsed twice as originating from the original article, suggesting that information from the Twitter feed may have been integrated into the memory of the article. For each participant, we calculated the proportion of suggested items reported as present in the article during the recognition test and reported as either “in the article only” or “in both [the article and Twitter feed] and they were the same” during the source test. Using this stronger measure of false memory, we found that participants who read the Iowa article showed higher false memory than those who read the MSU article, \( F(1, 152) = 8.45, p = 0.004, \eta_p^2 = 0.05 \) but there was no main effect of Twitter source, \( F(1, 152) = 0.04, p = 0.85 \). Interestingly, we found a significant interaction between the factors: participants in congruent article and Twitter source conditions (i.e., MSU article...
and MSU Twitter; Iowa article and Iowa Twitter) showed lower false memory than participants in incongruent conditions (i.e., MSU article with Iowa Twitter; Iowa article with MSU Twitter), $F (1, 152) = 5.76, p = 0.02$, $\eta^2_p = 0.04$, (see Fig. 1). This effect is the opposite direction of our hypothesis.

Next, we assessed reading times for each article to ensure that our participants (MSU students) did not spend more time reading the MSU article. However, participants who read the Iowa article showed significantly longer reading times ($\text{mean} = 183.62\text{s}$, $\text{SD} = 64.42\text{s}$) than those who read the MSU article ($\text{mean} = 159.67\text{s}$, $\text{SD} = 60.26\text{s}$), $t (152) = 2.40, p = 0.02, d = 0.38$. Thus, the false memory differences between the articles cannot be attributed to differences in encoding times.

Additionally, we investigated self-reported attention and trust towards the article and Twitter feeds. We did not find a significant effect for greater attention to the article by participants who read the MSU article than participants who read the Iowa article, $F (1, 152) = 1.89, p = 0.10$, $\eta^2_p = 0.02$, despite shorter reading times. Surprisingly, we found that participants who read the MSU Twitter feed reported significantly higher attention to the article (collapsed across article) compared to those who read the Iowa Twitter feed, $F (1, 152) = 3.96, p = 0.05$, $\eta^2_p = 0.03$. This result should be interpreted cautiously as we did not ask about attention to the article until after the test. There was no interaction between the factors, $F (1, 152) = 0.75, p = 0.39$. For attention to the Twitter feed, there were no significant main effects and no interaction between the factors, all $Fs < 1$ and $p's > 0.3$.

Self-reported trust for the article was similar across conditions with no main effect of article, $F (1, 152) = 0.14, p = 0.71$, or Twitter feed, $F (1, 152) = 1.45, p = 0.23$, and no interaction, $F (1, 152) = 2.56, p = 0.11$. Trust in the Twitter feed did not vary by article, $F (1, 152) = 0.14, p = 0.71$, or Twitter feed, $F (1, 152) = 1.45, p = 0.23$, and there was no significant interaction, $F (1, 152) = 1.29, p = 0.26$.

5. Discussion

Contrary to our original hypotheses, false memory did not vary by article or Twitter source. We hypothesized a main effect of article on false memory, with greater false memory for the less familiar Iowa article than the more familiar MSU article. However, we found no differences in false memory between the two articles. We further hypothesized that false memory would be greater in conditions with heightened perceived familiarity of the Twitter authors with their content. Such an effect would be supported by greater false memory in conditions for which the purported Twitter authors discussed information about their own university. The lack of significant interactions of article and Twitter source for false alarms to suggested information precludes us from directly supporting our second hypothesis.

However, in our more stringent false memory analysis that incorporated source memory results, we found a significant main effect of article that is consistent with our prediction and replicates earlier work (Prentice et al., 1997). Participants who read the less familiar (Iowa) article were more likely to incorrectly identify the information with the article than participants who read the more familiar (Iowa) article were more likely to incorrectly identify the information with the article by participants who read the less familiar (Iowa) article were more likely to incorrectly identify the information with the article. Authors less familiar with a topic might therefore be perceived as more reliant upon the article to gather information for their tweets. For example, if a psychologist and a journalist each provided information about the brain, an individual who receives this information might assume the journalist obtained their information from external sources, which are more credible, whereas the psychologist might rely on memory, which is fallible. Although counterintuitive, individuals might have greater skepticism of the accuracy of social media users who are more familiar with the topics they post online. The current data suggest that this might be possible, but future research and replication are required before we can make any conclusions based on this data.

This study is an important first step in understanding how familiarity affects false memory from social media. We hope that continued investigation will illuminate how individuals combine contextual information with source accuracy to find effective methods of reducing memory distortion. The expanding influence of social media in everyday life and its growing role in the dissemination of news about current events demands a greater understanding of the effects the Internet has on false memory formation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.chb.2017.08.032.

References


