Psychon Bull Rev (2014) 21:1551–1556
DOI 10.3758/s13423-014-0639-9

BRIEF REPORT

The effect of Twitter exposure on false memory formation

Kimberly M. Fenn · Nicholas R. Griffin · Mitchell G. Uitvlugt · Susan M. Ravizza

Published online: 14 May 2014

Abstract Social media sites such as Facebook and Twitter have increased drastically in popularity. However, information on these sites is not verified and may contain inaccuracies. It is well-established that false information encountered after an event can lead to memory distortion. Therefore, social media may be particularly harmful for autobiographical memory. Here, we tested the effect of Twitter on false memory. We presented participants with a series of images that depicted a story and then presented false information about the images in a scrolling feed that bore either a low or high resemblance to a Twitter feed. Confidence for correct information was similar across the groups, but confidence for suggested information was significantly lower when false information was presented in a Twitter format. We propose that individuals take into account the medium of the message when integrating information into memory.

Keywords Memory · False memory · Social media

The Internet has become a household and workplace necessity, and an estimated 80 % of Americans have Internet access (Zickuhr & Smith, 2012). Although the Internet has a multitude of uses (e.g., entertainment, communication), many use it to gather information. For example, more Americans obtain their daily news from online sources than from printed sources like newspapers, and Internet news sources are second only to television in popularity (Rosenstiel & Mitchell, 2011). Individuals seek information online not only from reputable news sites, but also from social media sites. An estimated 67 % of Internet users engage with some form of social media or social networking, such as Facebook, Twitter, or Pinterest (Duggan & Brenner, 2012).

Gathering information from social media sites is potentially problematic, however, because these sites may not always present accurate information. Instead, individuals are able to post freely about any topic without verification, thereby potentially permitting dissemination of inaccurate information. A vast literature has shown that false information given after an event can significantly alter the original memory (for reviews, see Ayers & Reder, 1998; Frenda, Nichols & Loftus, 2011; Loftus, 2005). In the same way, reading social media posts containing false information may change memory for information acquired from more reliable sources.

The present study assessed whether inaccurate information presented through social media affects false memory formation differently than inaccurate information acquired in a non-social-media format. There are various forms of social media, but we chose to use Twitter for the present experiment because we felt that it would be the most ecologically valid presentation format. Although Twitter, with 230 million monthly active users (Twitter.com, 2014), is not used as widely as Facebook, with 874 million monthly active users (Facebook.com, 2014), we felt that it would provide a more realistic way to investigate social media use. Facebook users typically know or have met their "friends" or, at least, have mutual acquaintances. In contrast, individuals frequently "follow" Twitter users whom they have never personally met. Thus, information acquired via Twitter could be compared with non-social-media sources without a potential confound of interpersonal association.

Two factors may modulate the formation of false memory when inaccurate information is presented via Twitter. First, it is possible that when false information is conveyed through Twitter, false memory may be higher than if the same

Electronic supplementary material The online version of this article (doi:10.3758/s13423-014-0639-9) contains supplementary material, which is available to authorized users.

K. M. Fenn (✉) · N. R. Griffin · M. G. Uitvlugt · S. M. Ravizza
Department of Psychology, Michigan State University, East Lansing, MI 38824, USA
e-mail: kfenn@msu.edu

information is presented in more formal sources (e.g., Internet news outlets, television, newspapers/magazines) because of its informal, conversational tone. Although false memory has not been studied in the context of social media, a recent study reported that memory for *correct information* was more accurate for Facebook posts than for book excerpts, headlines, or even comments about news articles (that were posted online). This effect was not due to deeper, more social encoding of information but was potentially due to the informal tone or gossipy nature of the selected posts, which mimicked spontaneous thoughts (Mickes et al., 2013). Therefore, false information acquired via Twitter might be particularly likely to be integrated into existing memory, resulting in higher false memory rates.

Another factor, however, may produce the opposite effect; namely, Twitter may have low credibility and decrease the rate of false memory formation. The credibility of a source has been shown to influence the formation of false memory; when participants are given false information by a source that is untrustworthy or has low credibility, false memory is reduced (Echterhoff, Hirst & Hussy, 2005). It is therefore possible that individuals are aware that Twitter can have low credibility, and information that is acquired via Twitter may be subject to higher scrutiny than information acquired via sources that are not associated with social media. If this were the case, we would expect that false information that is acquired from Twitter might be resistant to integration into existing memory representations.

Testing false memory formation with Twitter allows us to evaluate this process in unique ways. Although false memory studies are ubiquitous, none have tested the effects of conversational style. In fact, testing such variables might be difficult in traditional study formats where misinformation is presented verbally or in text narrations. In such a framework, informal language might be unusual, particularly within an experimental setting. In contrast, informal language is expected in social media formats. Thus, one aim of the present study was to investigate whether informal language affects false memory formation within the context of social media. Furthermore, the present study allowed us to directly investigate how differences in presentation format affect false memory. McLuhan (1964) famously proposed that the medium of presentation affects the message it conveys and that individuals approach messages differently when they are delivered through different types of media. In both our experimental and control conditions, false information was provided by the same source, whose credibility should be the same (i.e., participants in both conditions are told that the material was written by participants in a previous experiment). Thus, instead of testing the credibility of an *individual*, we tested the credibility of the *medium*. In other words, if the same people are presenting information, does the credibility of the format affect rates of false memory formation?

To investigate these two questions, we presented participants with a series of images that depicted a story and then presented false information either through a Twitter feed or through an information feed that was not associated with social media. The Twitter group was further divided such that the false information was presented using informal language for about half of the participants and more formal language for the other half. We first assessed whether informal language used on Twitter prompts higher rates of false memory, similar to the way correct memory was higher when it was presented in Facebook posts, as compared with book excerpts (Mickes et al., 2013). To answer this question, we compared memory rates for information given through Twitter using formal versus informal language. We then investigated whether false memory would be affected by the medium of presentation by comparing the Twitter condition with the non-social-media condition.

Method

Participants

One hundred seven native English-speaking undergraduate students (74 females, mean age = 19.3 years, *SD* = 1.69) from Michigan State University participated in the experiment for course credit. An additional 11 individuals completed the experiment but were excluded from all analyses either because they were not native English speakers (n = 3) or because of experimenter error during testing (n = 8). Seventy-three of the participants (68 %) reported having Twitter accounts, and 1 additional participant formerly had an account but had deleted it in recent months.

Materials and procedure

All participants completed three experimental phases: encoding, misinformation, and confidence test. During encoding, participants viewed a series of 50 images that depicted a story of a man robbing a car (adapted from Okado & Stark, 2005). Each picture was displayed for 5 s, with a 500-ms interstimulus interval. After encoding, participants completed the operation span (OSPAN) task (Unsworth, Heitz, Schrock & Engle, 2005) to reduce rehearsal of information. In the OSPAN task, participants were asked to maintain verbal information in memory while performing arithmetic problems. On each item, participants performed simple mathematical verifications (e.g., 6 * 3+1 = 19) and responded by clicking on "True" if the statement was correct or "False" if the statement was incorrect. After each problem, they were given a letter to remember (from a set of 12 consonants). After 3–7 problems and letters, they were asked to recall the letters in the order in which they had been presented. During the

recall phase, all 12 possible letters appeared in a grid on the screen, and participants clicked on the letters in the order in which they had been presented. After the letter recall phase, they were given feedback on the number of letters that were correctly recalled and the number of math errors for that set of trials. Participants completed three trials of each set size (3, 4, 5, 6, and 7 letters) and, in total, were asked to remember 75 letters amid 75 math problems.

During the misinformation phase, all participants viewed an information feed that presented 40 lines of text that narrated the events depicted in the images. While most of the information in the feeds was accurate, each participant was exposed to six details that directly conflicted with the images. For example, one line of text read, "The car had a Harvard sticker in the back window," when the car seen in the images actually had a Johns Hopkins sticker in the back window. Participants were not warned that they would see contradictory information.

To assess the effect of Twitter on false memory formation, participants were pseudorandomly assigned to one of three conditions: Twitter (n = 37), control (n = 40), or Twitter-control (n = 30). In all conditions, the information feeds were designed to provide information in a format similar to a Twitter feed. Each feed was a two-panel ticker in which new text appeared at the top. After 5 s, that text scrolled down to the lower panel, where it remained for an additional 5 s. Thus, each line of text was present for a total of 10 s.

The information feeds were visually similar; they were the same size and had the same format, but they differed in background designs across the conditions. In the Twitter and Twitter-control conditions, the information feed was labeled "Tweet Ticker" and was designed to resemble an official Twitter ticker; it had a light blue background and an image of the Twitter bird logo in the bottom right corner. Furthermore, a spatially blurred image of a person appeared to the left of the text, and black bars were placed at the top of each panel and scrolled with each response, creating the illusion of user censorship. The control feed was labeled "Photo Recap" and had a red background. Blurred images also appeared to the left of the text, but the feed did not contain censorship bars or social media logos (Fig. 1).

In the Twitter and Twitter-control groups, we told participants that the information feed consisted of tweets written by previous participants who had viewed the same images and tweeted responses as the images were presented. We told participants that the names and faces of the participants were blocked to protect their identities. To further strengthen our cover story, the text in the feed for the Twitter group was designed to resemble text found online; it was written with informal language and syntax. Moreover, some lines incorporated hashtags (#) or at signs (@), which are frequently used in tweets. Finally, to access the Twitter feed, experimenters navigated to a laboratory website and then clicked on a link to the Twitter feed. The control group was also told that the information in the feed was written by participants who had seen the photos in a previous experiment. This allowed us to isolate the effect of the medium of presentation, because all groups believed that the text had been written by past participants. The text for the control group was written in more formal language and consisted of complete sentences. To control for possible effects of language, the Twitter-control group was given the same cover story and visual display as the Twitter group but had the more formal text from the control condition (see Table 1, Appendix A, and Appendix B in the Supplemental Online Material [SOM]).

During the test phase, we presented events or details that may have appeared in the images and asked participants to respond on the basis of their memory for the images. Each stimulus was individually presented, and participants rated their confidence that the information was present in the images, using an 8-point Likert scale ranging from *definitely did not see in pictures* to *definitely saw in pictures*. There were 36 items in the test: information that appeared in the images *and* information feeds (n = 10), information that appeared only in the images (n = 10), information that appeared only in the information feeds (i.e., false information; n = 6), and novel lures that appeared neither in the images nor in the feeds (n = 10). For example, one item

Fig. 1 Examples of the information feeds in the Twitter (leftmost), control (center), and Twitter-control conditions. The text in the control and Twitter-control conditions was identical, with the exception of text following hashtags (#) and at signs (@). Note that the text appears slightly larger (proportionally) in these images to increase legibility

that appeared in both the images and the feeds was, "A man was walking down the street while wearing a Yankees t-shirt." The man's t-shirt was clearly visible in the images, and it was also mentioned in the first line of the scrolling feeds (see Appendix C in the SOM for a full list of test items). An example of an item that appeared only in the information feeds was, "The car had a Harvard sticker in the back window." No time limit was imposed on responses.

Directly after the confidence test, we asked participants to rate how much attention they had paid to the information feed and how much they trusted the information in the feed, both on 5-point Likert scales from *not at all* to *very much*.

Finally, participants completed a source test. Each item from the test (except for novel lures) was presented again, and we asked participants to report where they had obtained the information about that item. Five options were provided: "saw it in the pictures only," "saw it in the information feed only," "saw it in both [the pictures and information feed] and they were the same," "saw it in both and they were different," and "guessed." We were most interested in how often participants would attribute our suggested information to the images (the first and third options). This would constitute a richer form of false memory, since it would suggest some memory of the information appearing in the photos.

Results

Our Twitter group was divided such that half of the participants read tweets with informal language and half read tweets in more formal language. We first tested the effect of language in the Twitter feeds on memory performance by comparing correct and false memory in the Twitter and Twitter-control conditions. We compared average confidence ratings for correct information that appeared only in the images and information that appeared in both the images and the tweets. This analysis showed that language did not significantly affect confidence for correct items. For items that appeared in both the images and tweets, the group that read the Twitter feed in informal language (Twitter) showed mean confidence ratings of 6.92 ± 0.71 (*SD*), and the group that read the more formal tweets (Twitter-control) had average confidence ratings of 6.86 ± 0.69, $t(65) = 0.34$, $p = .73$. For information that appeared in the images only, the groups showed mean confidence of 5.75 ± 0.96 (informal) and 5.4 ± 0.96 (formal), $t(65) = 1.39$, $p = .16$. Next, we compared confidence for suggested information (from the Twitter feeds), and the two groups again showed similar performance (3.67 ± 1.5 and 3.54 ± 0.79, informal and formal language, respectively), $t(65) = 0.41$, $p = .67$. Finally, confidence for novel items was also similar in the two groups (1.9 ± 0.62 and 2.1 ± 0.76, informal and formal language, respectively), $t(65) = 0.96$, $p = .34$. Because the two Twitter groups did not differ on any of our memory measures, we collapsed across them for all subsequent analyses.

We first assessed confidence for correct information between the collapsed Twitter group and the control group. As can be seen in Fig. 2, the Twitter group showed similar confidence (6.89 ± 0.69) for information that appeared in both the images and the feed as the control group (6.69 ± 0.74), $t(105) = 1.40$, $p = .15$. The groups also showed similar confidence for information that appeared only in the images (5.6 ± 0.96 and 5.55 ± 1.04, Twitter and control, respectively), $t(105) = 0.27$, $p = .78$.

Our primary interest in this study was to investigate false recognition of suggested information. The Twitter group showed significantly lower confidence (3.61 ± 1.27) for suggested information than did the control group (4.24 ± 1.83), $t(105) = 2.06$, $p = .04$, $d = 0.39$ (Fig. 3). Finally, we analyzed confidence for novel lures and found that the two groups did not differ, $t(105) = 0.42$, $p = .66$. Thus, the groups differed only in their confidence for suggested information, which argues against any sort of global response bias.

In addition to basic confidence for the suggested information, we also assessed how often participants attributed their memory for suggested information to the images, as a stronger measure of memory distortion. For each participant, we calculated the proportion of items for which the participant reported high confidence (a rating of either 7 or 8), which would suggest that the participant responded consistently with the suggestion[1] *and* reported seeing the information *in the photographs* (in the source test). This stronger form of false memory was quite low for both the Twitter (2.7 % ± 8.0, mean ± *SD*) and the control (7.9 % ± 15) groups. Consistent with earlier results, the control group showed higher false memory than did the Twitter groups, $t(105) = 2.32$, $p = .02$, $d = 0.4$. We performed the same analysis for information that did appear in the images and found no difference between the groups. For information that appeared in both the images and the feeds, the Twitter (63.4 % ± 16.9) and control (63.5 ± 13.0) groups showed very similar response patterns, $t(105) = 0.02$, $p = .98$. Similarly, for information that appeared only in the pictures, the two groups did not differ reliably (Twitter, 49.3 ± 18.5; control, 49.7 ± 20.8), $t(105) = 0.12$, $p = .89$.

Previous studies have shown that working memory capacity (WMC) predicts false memory in the Deese–Roediger–McDermott paradigm (Peters, Jelicic, Verbeek & Merckelbach, 2007; Watson, Bunting, Poole & Conway, 2005) and the misinformation paradigm (Zhu et al., 2010).

[1] For this analysis, we first assessed responses on the confidence test. For each item, we used confidence ratings of either 7 or 8 as an indication of high confidence that the item had appeared in the images. We then calculated the number of times that the participant attributed these high confidence ratings to actually remembering the information from the images (by selecting either "I saw it in the images only" or "I saw it in the images and feed and it was the same" during the source test.

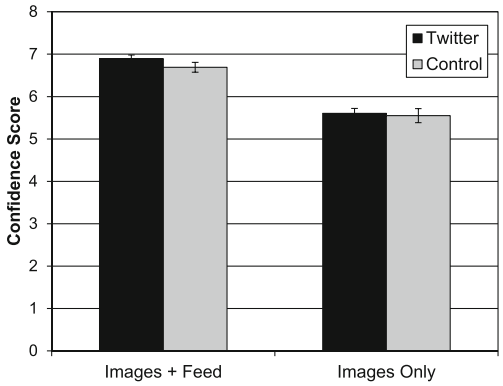

Fig. 2 Average confidence rating (±*SEM*) for correct information that appeared in the images and the information feeds (left bars) and information that appeared only in the images (right bars) for the Twitter and control conditions

To ensure that differences in WMC could not explain our results, we compared performance on the OSPAN task. The control group (59.4 ± 13.2) showed similar performance as the Twitter group (57.7 ± 13.3), $t(105) = 0.66$, $p = .50$. To further ensure that WMC could not explain the differences between the groups, we performed a hierarchical regression analysis on confidence for suggested items. We entered OSPAN score in step 1 and condition in step 2 and found that although OSPAN score accounted for a significant amount of variance, $\beta = -.26$, $t = 2.79$, $p < .01$, condition accounted for an additional 4.6 % of the variance, $\beta = -.22$, $t = 2.33$, $p = .02$.

There are several reasons why participants might show lower confidence for suggested items in the Twitter condition. The control participants may have paid better attention to the information feed, or they may have trusted the information more than participants in the Twitter conditions. To examine this, we compared self-reported ratings of attention and trust. Ratings of attention did not differ between the Twitter (4.52 ± 0.78) and control (4.65 ± 0.73) groups, $t(105) = 0.83$, $p = .41$. However, there was a strong trend for the Twitter group to report trusting the information (2.73 ± 0.96) less than did the control group (3.15 ± 1.25), $t(105) = 1.94$, $p = .054$, but this effect narrowly missed significance. Thus, although both groups reported paying close attention to the information in the feed, there was a trend for them to trust the information less if it was acquired via Twitter.

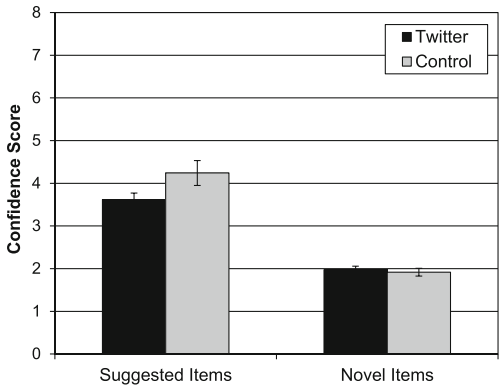

Fig. 3 Average confidence rating (±SEM) for incorrect information that was suggested in the information feeds (left bars) and for novel lures (right bars) for the Twitter and control conditions

Discussion

When conflicting information was presented using a Twitter feed, participants showed lower confidence for false information than when the same information was given using a non-social-media source. This effect is particularly powerful because information in both of the sources was ostensibly written by other participants. Thus, the source of the information was held constant across groups, so the present findings show that false memory not only depends upon the credibility of the source, but also depends upon the credibility of the medium of presentation. This is consistent with McLuhan's (1964) proposal that the medium of presentation affects the interpretation of the message it conveys. Importantly, confidence for valid information that appeared in the images and confidence for novel lures did not differ between the groups, so our results cannot be explained by simple response bias. Instead, we propose that individuals take into account the medium when integrating information into memory. Participants in the Twitter conditions reported trusting the information marginally less than did participants in the control condition. This pattern of results is consistent with previous work that has shown that when participants are given false information by a source that is untrustworthy or has low credibility, false memory is reduced (Echterhoff et al., 2005).

It is also possible that our results can be explained through reduced attention in the Twitter group or reduced motivation to encode information into long-term memory. Participants may have simply not attended to the Twitter feed as closely as the control feed, or they may have altered their memory strategies and reduced the amount of information that they attempted to encode into long-term memory. A recent study found that participants showed lower memory retention for information that they believed was saved on a computer than for information they believed had been erased (Sparrow, Liu & Wegner, 2011). This suggests that memory formation may be mediated by the future availability of information. Thus, when individuals encounter information via Twitter, they may alter their memory strategies and reduce the amount of information that they attempt to encode into long-term memory. Subjective ratings of attention did not differ between the groups, but this may have been due to demand characteristics. However, we do not believe that reduced attention or reduced motivation to remember the information can adequately explain all of our results, because if this were the case, we would expect to see lower confidence for correct information that appeared in both the images and the information feeds in the

Twitter condition than in the control condition. There was no difference between the groups, and if anything, the trend was in the opposite direction. Thus, we do not believe that the lower false memory could be due to a failure to encode the information in the Twitter condition.

The present study was designed to also investigate the effect of informal language on false memory from Twitter. Previous research found that the informal, conversational tone of Facebook posts led to increased correct memory for information (Mickes et al., 2013). However, we found that confidence for false information in the Twitter feed was similar regardless of whether information was presented conversationally or more formally. This may seem inconsistent with prior work, but the materials in the Mickes et al. study were quite different between conditions. They compared actual posts from Facebook with actual sentences from books, headlines, CNN articles, and even comments about the news articles that were posted by readers. Thus, the information that participants were asked to remember differed across conditions. Here, the basic information and content were held constant; only the language in the tweets varied. Importantly, these results suggest that the reduced confidence in false information in the Twitter group cannot be due to the informal language in Twitter. Instead, we propose that the critical factor in our study was the medium of presentation.

It is unclear whether our results would extend to other forms of social media or to more verified sources on Twitter. Twitter is unique in that individuals often receive information from other individuals whom they have never met. In contrast, if the same information were presented in a Facebook feed, by a friend or an acquaintance, we might obtain an opposite finding. Individuals may trust a source more if it is personally known to them, and there may, therefore, be higher rates of false memory if inaccurate information is presented through Facebook. Similarly, although many people use Twitter for personal use, there is also a growing network of professional Twitter sources (e.g., news outlets, politicians, etc.). If false information is presented on Twitter through an official source, participants may be more likely to accept the information and show higher rates of false memory. Finally, these results may be specific to young adults (ages 18–29), who are more familiar with and more likely to use social networking sites than are other age groups (Duggan & Brenner, 2012).

In conclusion, false information that was acquired via Twitter was less likely to be integrated into a memory representation. This effect is unlikely to have been due to differences in attention, since participants in both groups reported similar attention to the information and showed similar memory for information that was presented in both the images and the information feeds. Instead, individuals seem to take into account the medium of presentation when evaluating new information. We propose that young adults take into account *how* information is presented and not just *who* is presenting it.

Acknowledgements This study was funded by the National Science Foundation Early Development CAREER award (#1149078) to S.R.

References

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5*(1), 1–21.

Duggan, M., & Brenner, J. (2012). *The demographics of social media users - 2012*. Pew Internet & American Life Project, Pew Research Center. Retrieved from http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_SocialMediaUsers.pdf

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition, 33*(5), 770–782.

Facebook.com. (2014). Key facts [Website information]. Retrieved January 21, 2014, from newsroom.fb.com/Key-Facts.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science, 20*(1), 20–23.

Loftus, E. F. (2005). Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361–366.

McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.

Mickes, L., Darby, R. S., Hwe, V., Bajic, D., Warker, J. A., Harris, C. R., et al. (2013). Major memory for microblogs. *Memory & Cognition, 41*(4), 481–489.

Okado, Y., & Stark, C. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory, 12*(1), 3–11.

Peters, M. J. V., Jelicic, J., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology, 19*(2), 213–232.

Rosenstiel, T., & Mitchell, A. (2011). *Overview: The State of the News Media 2011*. Pew Project for Excellence in Journalism 2011, Pew Research Center. Retrieved from http://stateofthemedia.org/2011/overview-2/

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778.

Twitter.com. (2014). About Twitter, Inc [Website information]. Retrieved January 21, 2014, from about.twitter.com/company.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505.

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 31*(1), 76–85.

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., … Dong, Q. (2010). Individual differences in false memory from misinformation: cognitive factors. *Memory, 18*(5), 543–555.

Zickuhr, K., & Smith, A. (2012). *Digitial differences*. Pew Internet & American Life Project, Pew Research Center. Retrieved from http://pewinternet.org/Reports/2012/Digital-differences.aspx