

Detailed and gist-like visual memories are forgotten at similar rates over the course of a week

Nora Andermane · Jeffrey S. Bowers

Published online: 27 January 2015
© Psychonomic Society, Inc. 2015

Abstract A number of recent studies have highlighted the exceptional capacity and fidelity of visual long-term memory. For instance, Brady, Konkle, Alvarez, and Oliva (*Proceedings of the National Academy of Sciences*, 105, 14325–14329, 2008) presented participants with thousands of images for nearly 6 h and then tested their memory in a two-alternative forced choice (2AFC) task. Participants were 87% accurate, even when the foil was extremely similar to the target (e.g., when the same object was presented in a different state). In the present investigation, we extended these findings by including a one-week delay condition and by testing memory in a yes–no as well as a 2AFC task. We replicated the exceptional memory results at a short delay. However, following a week delay, recognition accuracy was greatly reduced in both tasks, with comparable reductions in performance when the foils were both similar and dissimilar. These findings suggest that detailed and gist-like visual memories decay at similar rates, which highlights important limitations of visual long-term memory.

Keywords Visual long-term memory · Exceptional memory · Recognition memory

Visual long-term memory is exceptional under some conditions. For example, Shepard (1967) found that participants were 98% accurate in a two-alternative forced choice (2AFC) task after studying a series of 600 pictures. These findings were extended in a landmark study by Standing (1973); he found that after viewing up to 10,000 images over

the course of several hours, people were able to choose the images they had seen in a 2AFC task with an accuracy of over 80%. Standing concluded that the capacity of visual memory for recognizing pictorial content is almost limitless. In addition, some of these early studies measured the longevity of visual long-term memory by including various study–test delay conditions, and memory performance was still excellent at the longer delays. For example, Shepard found that recognition performance was at ceiling (99.7%) following a 2-h study–test delay, and still remarkably accurate (87.0%) following a seven-day delay. Similarly, Nickerson (1968) explored visual recognition memory in a 2AFC task with delay conditions ranging from one day to one year, and concluded that visual memory retention is substantial, since the probability of correctly recognizing a seen image was approximately 90% after a week delay, and still higher than 70% after a month.

These findings show that visual long-term memory has the capacity to store and retrieve a vast number of images even after long delays. However, virtually all of these studies had employed a 2AFC task in which the target was paired with an unrelated foil image. Therefore, it is difficult to determine whether the retained memory representations contained gist-like information about the basic-level object category of the studied image (e.g., “I saw an elephant rather than a chair”), or whether the memories included high-fidelity information about the perceptual details of the studied image (e.g., “I saw this particular image of an elephant”).

Recent studies by Brady and colleagues (Brady, Konkle, Alvarez, & Oliva, 2008; Konkle, Brady, Alvarez, & Oliva, 2010a, b) have shed light on this issue by showing that people can recognize images with a remarkable level of detail. For example, in the study by Brady et al. (2008), participants viewed images of 2,500 objects for 5.5 h. Shortly after the study phase, they were given a 2AFC task in which a previously studied item was paired with a foil, and participants had to identify the studied image. The study was based on three

This research was supported by Leverhulme Grant RJ5538 awarded to Jeffrey S. Bowers

N. Andermane · J. S. Bowers (✉)
School of Experimental Psychology, University of Bristol,
BS8-1TU Bristol, UK
e-mail: j.bowers@bristol.ac.uk

types of foils with varying levels of similarity to the target: unrelated objects (e.g., a skirt and a jar), different exemplars from the same basic object category (e.g., a light starfish and a dark starfish), and images of the target object in a different state (e.g., the same abacus with beads in different configurations). See Fig. 1 for some examples. Brady and colleagues found that participants were able to identify target images with an accuracy of 93% when the foil was unrelated to the target, 88% when the foil was from the same object category, and 87% when the foil was the same object but in a different state. These findings show that the long-term visual memory representations stored were highly detailed.

However, there has been a missing link between the early studies highlighting the longevity of visual memories and the more recent work concerning the high fidelity of these memories. That is, recognition memory for high-fidelity visual content has only been assessed following a short study–test delay, and accordingly, it is possible that this highly detailed information would decay or become less accessible over the course of longer delays. Indeed, a variety of empirical evidence has suggested that detailed memories fade more quickly than more general or gist-like memories (e.g., Lampinen, Copeland, & Neuschatz, 2001; Tuckey & Brewer, 2003), and these observations have been embedded in theories such as fuzzy-trace theory (Brainerd & Reyna, 2002). Accordingly, it might be expected that memory performance would decline more quickly when the memory foils were similar than when they were less similar.

In the present study, we directly assessed the longevity of visual long-term memory over a week when the foils at test were similar and dissimilar, using both the 2AFC and yes–no recognition tasks. We included the yes–no task because it is arguably more ecologically valid than the 2AFC task (we are

rarely confronted with the same object in two different states, presented side by side), and because overall performance, in terms of percentages correct, is reduced in the yes–no relative to the 2AFC task (see, e.g., Macmillan & Creelman, 1991). Accordingly, it was of interest to directly assess exceptional memory performance in these two tasks.

Method

Participants

Thirty-two psychology undergraduate students from the University of Bristol (18–25 years of age) took part in this study. All of the participants had normal or corrected-to-normal vision and received course credit for their time.

Stimuli

Digital color photographs depicting animals and objects were collected from Brady et al. (2008). One hundred pairs of images depicted categorically distinct objects (*novel* pairs), 100 pairs depicted different exemplars from the same object category (*exemplar* pairs), and 100 pairs depicted the same object in two different states (*state* pairs). An additional 1,200 filler images were also selected. The images (256 × 256 pixels) were presented in the center of a white screen on a Viglen desktop computer using the DMDX experiment software (Forster & Forster, 2003). The images were enlarged to the size of approximately 8 × 8 cm.



Fig. 1 Examples of target and foil pairs in the three foil conditions: novel, exemplar, and state. Adapted from “Visual Long-Term Memory Has a Massive Storage Capacity for Object Details,” by T.F. Brady, T.

Konkle, G.A. Alvarez, and A. Oliva, 2008, *Proceedings of the National Academy of Sciences*, 105, p.14326. Freely available online through the PNAS open access option

Design and procedure

Half of the participants ($N = 16$) completed the test phase 10 min after completion of the study phase (Day 1), whereas the other half completed the test phase seven days after the study phase (Day 8). The same computer was used in the study and test phases in all cases. Participants were instructed to closely attend to each image presented during study because their memory would be assessed at a following test phase.

The study phase took approximately 2.5 h to complete and consisted of 1,500 trials. Participants were informed that some of the images would repeat and their task was to detect these repeating images by pressing a key, and that no response was required for nonrepeating images. On each trial, a fixation point was presented in the center of the screen for 800 ms, followed by an image presented for 3 s, and depending on the response, feedback text appeared for 500 ms: “Hit” after a correct detection of a repeating image, “False alarm” after a response to a nonrepeating image, “Miss” after failing to detect a repeating image, and no feedback if no response was made to a nonrepeating image.

The repeat detection task was introduced in order to ensure that participants would pay attention to each image during study phase, and it resembled the filler task used by Brady et al. (2008). Of the 1,200 filler images, 200 were repeated overall; 40 of the images were repeated at the following intervals: 1, 5, 10, 50, and 100 intervening trials. The target images and nonrepeating filler images were presented randomly among the repeating filler images, and no target image was repeated. All of the images were presented in six blocks (250 images in each block), each block lasting ~20 min. Participants were encouraged to take a 5-min break between blocks. After completing the study phase, the Day 1 group had a break for 10 min before moving on to the test phase, whereas the Day 8 group came back after 7 days to complete the test phase.

The test phase took about 30 min to complete. Participants were instructed that their memory would be assessed with a 2AFC and a yes–no task, and that both tests would include foil images that would look very similar to the images they had actually seen. Participants were also told that no feedback would be given during the test phase and that accuracy was more important than speed, so they should take their time and respond carefully. Of the 300 target images presented in the study phase, 150 images were presented in the 2AFC task and 150 were presented in the yes–no task. All images were presented in random order, with the two tasks intermixed. The pairs of images in the novel, exemplar, and state conditions were counterbalanced across participants, so that each member of the pair was sometimes the target in both the 2AFC and yes–no tasks, and so that each image was presented on both the left and the right in the 2AFC task.

Results

Repetition detection task

The overall performance of the Day 1 group was excellent (91.0% hits and 0.8% false alarms), and their memory was a function of delay, with 95.3%, 93.9%, 93.8%, 89.1%, and 83.0% hits following intervals of 1, 5, 10, 50, and 100 intervening trials, respectively. Similar results were obtained in the Day 8 group, with overall excellent performance (87.9% hits and 1.1% false alarms) and memory declining as a function of delay, with 94.5%, 91.2%, 91.3%, 83.2%, and 79.3% accuracy following 1, 5, 10, 50, and 100 intervening trials. These results are similar to those from the repetition detection filler tasks used in previous studies (e.g., Brady et al., 2008; Konkle et al., 2010a, b), and highlights that participants paid careful attention to the images during the study phase.

Recognition memory tasks

A longstanding debate has concerned whether the 2AFC and yes–no tasks are supported by the same or by different underlying memory processes (e.g., Bayley, Wixted, Hopkins, & Squire, 2008; Migo, Montaldi, Norman, Quamme, & Mayes, 2009), and a related controversy concerns the correct method of computing d' in order to compare memory performance across the two tasks (given that different signal detection models are associated with different theoretical claims about memory; Jang, Wixted, & Huber, 2009). For our purposes, this debate was not critical. Rather, the key question was whether exceptional memory performance would extend over the course of a week in these two tasks, and whether the high-fidelity memories required to distinguish targets from similar foils would be disproportionately lost over time, consistent with the claim that gist-like memories last longer than detailed ones. Accordingly, in the analyses below we analyzed the percentage accuracy scores of the two tasks separately, although we report standard measures of d' in both tasks, as well.

In Fig. 2, we report the mean accuracies in the 2AFC task in the novel, exemplar, and state conditions on Days 1 and 8, with associated d' values listed in parentheses. The d' values were obtained using the formula $d' = [z(H) - z(F)]/\sqrt{2}$, where H = hits, F = false alarms, and z = z value (Macmillan & Creelman, 1991). As expected, performance on Day 1 was excellent, and recognition accuracy was highest in the novel foil condition (90.0%) and somewhat reduced in the exemplar (85.0%) and state (82.1%) conditions. On Day 8, memory performance was greatly reduced, although still well above chance, with accuracies of 70.8% in the novel, 65.6% in the exemplar, and 62.3% in the state conditions, respectively. A 2 (delay) \times 3 (foil condition) mixed-samples analysis of variance (ANOVA) on the accuracy results revealed a significant

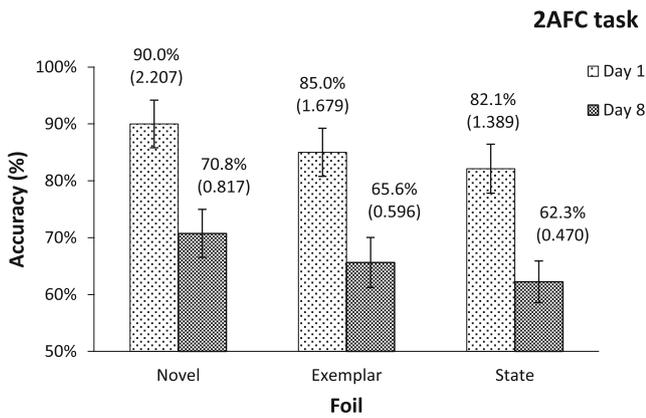


Fig. 2 Mean recognition accuracy (as percentages) in the two-alternative forced choice task in the novel, exemplar, and state foil conditions on Days 1 and 8. Error bars represent the 95% confidence intervals, and the d' values per each condition are indicated in parentheses

effect of delay, $F(1, 31) = 75.12, p < .001$, partial eta-squared (η_p^2) = .72, a significant effect of foil condition, $F(2, 30) = 17.52, p < .001, \eta_p^2 = .37$, and critically, no interaction between delay group and foil condition, $F(2, 30) = .03, p = .97, \eta_p^2 = .001$. Paired-samples two-tailed t tests revealed that on Day 1, accuracy was significantly higher in the novel than in either the exemplar, $t(15) = 2.89, p = .01, d = 0.63$, or the state, $t(15) = 4.18, p = .001, d = 0.99$, foil condition. No significant differences emerged between the state and exemplar conditions, $t(15) = 1.39, p = .18, d = 0.36$. The same pattern of differences between foil conditions was obtained on Day 8, with higher accuracy in the novel than in the exemplar, $t(15) = 2.34, p = .03, d = 0.63$, and state, $t(15) = 4.98, p < .001, d = 1.14$, conditions, and no significant differences between the exemplar and state conditions, $t(15) = 1.53, p = .15, d = 0.44$.

In Fig. 3, we report the mean accuracies in the yes–no task in the novel, exemplar, and state conditions on Days 1 and 8, with associated d' values listed in parentheses. Accuracy

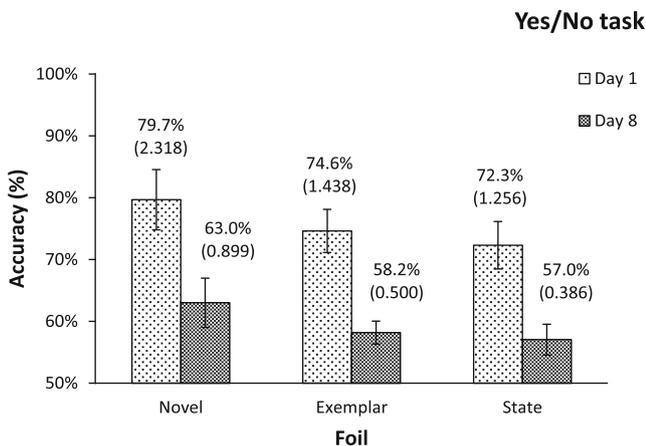


Fig. 3 Mean recognition accuracy (as percentages) in the yes–no recognition task in the novel, exemplar, and state foil conditions on Days 1 and 8. Error bars represent the 95% confidence intervals, and the d' values per each condition are indicated in parentheses

percentages were computed by adding correct identifications and correct rejections and dividing this sum by the total number of trials. The d' values were calculated according to the formula $d' = z(H) - z(F)$ (Macmillan & Creelman, 1991).

As can be seen from the figure, the pattern of accuracy results was similar to that from the 2AFC task across conditions, with memory performance that was reduced overall, relative to the 2AFC task. On Day 1, performance was best with novel foils (overall accuracy = 79.7%; hits = 63.6%, false alarms = 4.3%), was lower with exemplar foils (overall accuracy = 74.6%; hits = 65.9%, false alarms = 16.6%), and was lowest in the state condition (overall accuracy = 72.3%; hits = 69.5%, false alarms = 24.9%). Memory substantially declined on Day 8, with overall accuracy rates of 63.0% in the novel (hits = 39.0%, false alarms = 13.0%), 58.2% in the exemplar (hits = 38.8%, false alarms = 22.4%), and 57.0% in the state (hits = 41.7%, false alarms = 27.6%) conditions. A 2 (delay) \times 3 (foil condition) mixed-samples ANOVA showed a significant effect of delay, $F(1, 31) = 67.89, p < .001, \eta_p^2 = .69$, a significant effect of foil condition, $F(2, 30) = 18.29, p < .001, \eta_p^2 = .38$, and critically, no interaction between delay and foil condition, $F(2, 30) = 0.22, p = .81, \eta_p^2 = .01$. On Day 1, accuracy was highest in the novel foil condition, and significantly lower in both the exemplar, $t(15) = 2.83, p = .01, d = 0.64$, and state, $t(15) = 4.60, p < .001, d = 0.90$, conditions; however, we observed no significant differences between the state and exemplar conditions, $t(15) = 1.41, p = .18, d = 0.34$. This pattern was also present on Day 8 [novel vs. exemplar, $t(15) = 2.65, p = .02, d = 0.83$; novel vs. state, $t(15) = 3.35, p = .004, d = 0.96$; exemplar vs. state, $t(15) = 1.30, p = .21, d = 0.28$].

Discussion

Consistent with Brady et al. (2008)’s results, our participants showed exceptional visual recognition memory performance in a 2AFC task following a brief study–test delay, with 90% accuracy on trials in which the foils were unrelated, and over 80% accuracy when the foils were highly similar. However, memory performance was greatly reduced after one week, with performance reductions of approximately 20% in the 2AFC and 15% in the yes–no task. Furthermore, recognition memory performance was about 10% lower in the yes–no than in the 2AFC task across all task conditions. In the most difficult condition, in which memory was tested with similar foils in a yes–no task following a week delay, performance was not much above chance, falling to 57%.

Our key finding, however, is that the rates of forgetting were similar, whether the foils were similar or dissimilar to the targets, suggesting that detailed and gist-like visual memories were lost at similar rates. This appears to conflict with

the findings from a variety of studies (e.g., Lampinen et al., 2001; Tuckey & Brewer, 2003) that have reported faster forgetting for detailed visual memories. Our findings also appear to provide a challenge to the fuzzy-trace theory, according to which detailed (or *verbatim*) memories are forgotten more quickly than gist memories. The faster forgetting for detailed information is how fuzzy-trace theory accommodates the increased rate of false memories over time in many contexts (Brainerd & Reyna, 2002). Nevertheless, a slightly modified version of the theory might help reconcile our findings with past work.

A central tenet of the fuzzy-trace theory is that detailed and gist memories are encoded, stored, and retrieved separately, and in addition, the relative roles of detailed and gist traces in supporting performance can be manipulated in a number of ways. For instance, the repeated presentation of target items at study would tend to increase the role of detailed memories, whereas presenting a set of words that are all associated with a nonpresented target (as in the Deese/Roediger–McDermott false memory paradigm; Deese, 1959; Roediger & McDermott, 1995) would tend to increase the role of gist memories. Nevertheless, the fuzzy-trace theory assumes that the rate of forgetting is always faster for detailed than for gist information. It is this latter claim that is difficult to reconcile with our present findings. However, if the rates of forgetting for detailed and gist memories depend on the quality of the stored detailed and gist memories, then the past and present findings regarding forgetting rates may be reconciled.

Consider the following differences between the studies that have compared forgetting rates for detailed and gist memories. In the present study, the encoding of detailed memory traces was favored, given that the to-be-remembered materials were colored photographs of single objects presented on white background, decoupled from any scenery or contextual elements that could contribute to expectations based on gist. Each image was presented for 3 s, and the task instructions emphasized studying each image carefully, since highly similar foils would be presented at test. In addition, the study phase was composed of a list of unrelated images. Taken together, these task conditions encouraged participants to encode the images of individual objects in detail, such that schemas and expectations could play relatively little role in contributing to gist memories. By contrast, in past studies that had reported faster forgetting rates for detailed memories, the encoding of gist memories was favored. These past studies have included studies that assessed memory for verbal materials (which, by their nature, include less-detailed perceptual information that can be used to support memory; e.g., Reyna & Kiernan, 1994, 1995) and studies that assessed memory for visual images presented in the context of meaningful scenes that could support gist-like inferences and expectations (e.g., Lampinen et al., 2001; Tuckey & Brewer, 2003). In addition, in the present study we included a test condition that is thought

to favor the retrieval of detailed memories (e.g., Brainerd & Reyna, 2002); namely, identical images were repeated at study and test. By contrast, previous studies that compared the forgetting rates for detailed and gist visual memories had used different encoding and retrieval contexts, such as in verbal cued-recall tests (e.g., Tuckey & Brewer, 2003), or different views of the objects at study and test (e.g., Lampinen et al., 2001). The fact that past studies have included study and test conditions that favored the role of gist may help explain the common conclusion that detailed memories are forgotten more quickly.

Indeed, a recent article by Guerin, Robbins, Gilmore, and Schacter (2012) has provided evidence that many past studies have overestimated the rate at which detailed memory traces are lost. Participants studied a list of objects (e.g., a particular image of an anchor), and then at test were presented with highly similar foil images (e.g., a different image of an anchor) in two conditions: namely, conditions in which the foil was paired with an unrelated image (and the correct response was to reject all the images) or with the target itself. When the foil was paired with unrelated images, participants often selected the foil. The high false alarm rate suggests that the detailed visual traces that distinguished the target from the foil had been forgotten. By contrast, when the foil was paired with the target, participants were highly accurate and rarely false alarmed. This shows that detailed memory traces were in fact stored, and the poor performance when foils were paired with unrelated images reflected a failure to use this information. The authors suggested that detailed memory traces are often inaccessible when different images are presented at study and test. This conclusion may apply to past studies that have reported fast forgetting of visual details, given that they did not repeat the same images at study and test.

In summary, our ability to store and retrieve massive amounts of high-fidelity information in the visual long-term memory system is striking, but these results are most impressive in the 2AFC task, and the effects are restricted to short delays. These constraints on exceptional memory should be highlighted, as well. Critically, under conditions that support the encoding and retrieval of detailed visual memories, detailed and gist information are forgotten at similar rates.

References

- Bayley, P. J., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2008). Yes/no recognition, forced-choice recognition, and the human hippocampus. *Journal of Cognitive Neuroscience*, *20*, 505–512. doi:10.1162/jocn.2008.20038
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*, 14325–14329. doi:10.1073/pnas.0803390105

- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164–169. doi:10.1111/1467-8721.00192
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17–22. doi:10.1037/h0046671
- Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116–124. doi:10.3758/BF03195503
- Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68–78.
- Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General*, *138*, 291–306. doi:10.1037/a0015525
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*, 558–578. doi:10.1037/a0019165
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*, 1551–1556. doi:10.1177/0956797610385359
- Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1211–1222. doi:10.1037/0278-7393.27.5.1211
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.
- Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology*, *62*, 1198–1215.
- Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science*, *11*, 58.
- Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology*, *30*, 178–191.
- Reyna, V. F., & Kiernan, B. (1995). Children's memory and metaphorical interpretation. *Metaphor and Symbol*, *10*, 309–331.
- Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories—Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803–814. doi:10.1037/0278-7393.21.4.803
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163. doi:10.1016/S0022-5371(67)80067-7
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222. doi:10.1080/14640747308400340
- Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*, 101–118. doi:10.1037/1076-898X.9.2.101