

Does visual expertise improve visual recognition memory?

Karla K. Evans · Michael A. Cohen ·
Rosemary Tambouret · Todd Horowitz ·
Erica Kreindel · Jeremy M. Wolfe

© Psychonomic Society, Inc. 2010

Abstract In general, humans have impressive recognition memory for previously viewed pictures. Many people spend years becoming experts in highly specialized image sets. For example, cytologists are experts at searching micrographs filled with potentially cancerous cells and radiologists are expert at searching mammograms for indications of cancer. Do these experts develop robust visual long-term memory for their domain of expertise? If so, is this expertise specific to the trained image class, or do such experts possess generally superior visual memory? We tested recognition memory of cytologists, radiologists, and controls with no medical experience for three visual stimulus classes: isolated objects, scenes, and mammograms or micrographs. Experts were better than control observers at recognizing

images from their domain, but their memory for those images was not particularly good ($D' \sim 1.0$) and was much worse than memory for objects or scenes ($D' > 2.0$). Furthermore, experts were not better at recognizing scenes or isolated objects than control observers.

Keywords Visual recognition memory · Expertise · Medical images · Medical experts · Real world scenes · Objects

Introduction

Studies of visual memory show that humans have an astonishing and robust ability to recognize a vast number of previously viewed scenes and/or objects (Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989; Shepard, 1967; Standing, 1973). Shepard (1967) showed observers 612 different pictures, and obtained 98% correct recognition performance. Standing (1973) went even further, presenting his observers with 10,000 different photographs for 5 s each, and found 83% correct recognition performance. This type of recognition memory is far superior to verbal memory (Standing, Conezio, & Haber, 1970) and can last for at least a week (Dallet, Wilcox, & D'Andrea, 1968). More recently, Brady, Konkle, Alvarez, and Oliva (2008) showed that substantial detail is represented in this massive memory. Their observers studied 2,500 images, and were 88% correct even when correct performance required memory for which exemplar they had seen (e.g., this backpack, not that one). In addition, they were 87% correct when they had to remember the state or pose of images (e.g., the open backpack, not the closed one). In a different demonstration of the importance of detail, Vogt and Magnussen (2007b) showed that memory performance dropped from 85 to 65% correct for images of same picture category (e.g., colored picture of doors) once

K. K. Evans (✉) · T. Horowitz · J. M. Wolfe
Harvard Medical School, Visual Attention Lab,
64 Sidney Street, Suite 170,
Cambridge, MA 02139, USA
e-mail: kevens@search.bwh.harvard.edu

R. Tambouret
Harvard Medical School,
64 Sidney Street, Suite 170,
Cambridge, MA 02139, USA

K. K. Evans · T. Horowitz · E. Kreindel · J. M. Wolfe
Brigham and Women's Hospital,
Boston, MA, USA

R. Tambouret
Massachusetts General Hospital,
Boston, MA, USA

M. A. Cohen
Harvard University,
Cambridge, MA, USA

details were removed even if the gist (e.g., “big red door”) was unchanged.

These studies used non-specialist observers (e.g., college students) memorizing images that might be incidentally encountered in everyday life (e.g., pictures from magazines or pictures of backpacks). However, many people develop extensive expertise with particular image classes which are not encountered in everyday life. For example, cytologists spend thousands of hours scrutinizing micrographs of cervical cells, while radiologists devote similar time to searching mammograms. Is this visual expertise accompanied by improved visual recognition memory? If so, is this effect limited to images within the domain of expertise, or do these experts learn to remember visual stimuli better in general? The answers can tell us something about the effects of training on visual recognition memory. In different domains of expertise like master chess players,

soccer or basketball players there is evidence that experts are better than non-experts at memorizing meaningful material from their domain of expertise (Allard, Graham, & Paarsalu, 1980; Chase & Simon, 1973; Frey & Adesman, 1976; McKeithen, ReitmanHenry, Judith, & Hirtle, 1981). Our purpose was to place expert visual memory in the context of the massive memory for scenes and objects found in non-expert populations.

Methods

Participants

The study had four groups of participants: cytologists [including both cytopathologists (MDs) and cytotechnologists (not MDs)]; radiologists specializing in mammog-

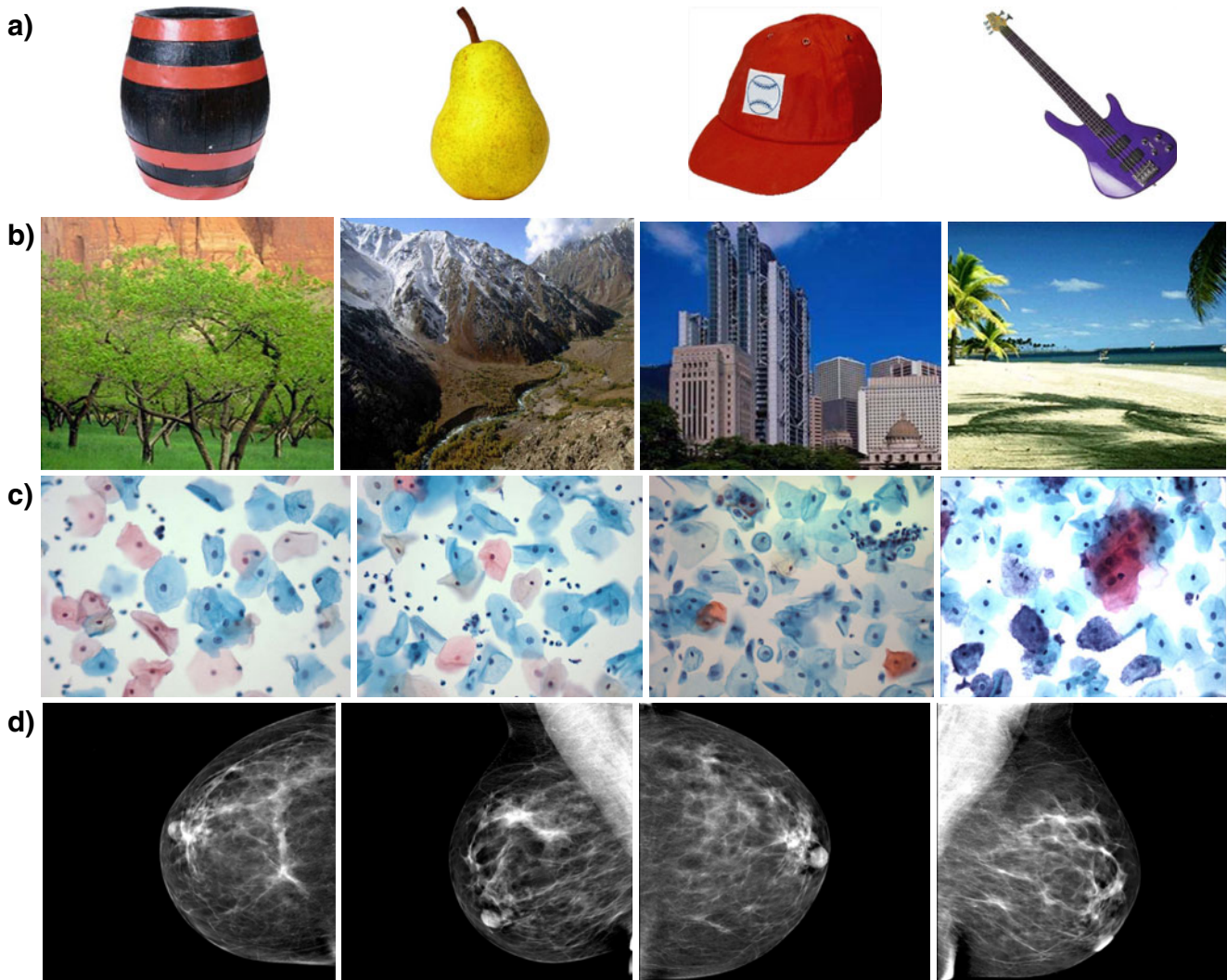


Fig. 1 Examples of stimuli used in different blocks of the experiment. **a** Objects, **b** scenes, **c** Pap smears, **d** mammograms. Cytologists and their control group were tested on **a**, **b** and **c**, while radiologists and their control group were tested on **a**, **b** and **d**

raphy; and two medically untrained control groups. The experimental cytology group consisted of 10 cytologists (6 females), age range 28–68 years with an average of 23 years of experience and a range of 1,500–10,000 cases diagnosed a year. The experimental radiology group consisted of 13 radiologists (12 females), age range 36–70 years with an average of 20 years of experience and a range of 1,000–8,000 cases diagnosed a year. The two control groups had 23 observers (11 females) with no medical background, age range 21–55 years. Each pair of groups (experimental and its control group) completed the same conditions and saw the same image sets. Each observer passed the Ishihara test for color blindness and had normal or corrected to normal vision. All observers gave informed consent, as approved by the Partners Healthcare Corporation IRB, and were compensated for their time.

Stimuli and apparatus

Each group saw three stimulus sets illustrated in Fig. 1: images of a variety of objects isolated on a white background (Fig. 1a); indoor and outdoor scenes (Fig. 1b); medical images, either “Pap smears” (i. e., cervical micrographs; Fig. 1c) or mammograms (breast X-rays; Fig. 1d). Each set contained 108 images. The images of scenes and objects were obtained from a public image dataset hosted by the Computational Visual Cognition Laboratory at MIT (<http://cvcl.mit.edu>). The Pap smear images were acquired by Rosemary Tambouret from her clinical practice. The mammograms were obtained from the Lee Bell Labs Breast Imaging center at Brigham and Women’s Hospital. The Pap smears and mammograms comprised an equal number of images with and without abnormalities. All the medical images were anonymized before the use in the experiment.

The experiment was conducted on a Macintosh computer running MacOS X, controlled by Matlab 7.5.0 and the Psychophysics Toolbox, version 3 (Brainard, 1997; Pelli, 1997).

Procedure

The experiment was composed of study and test phases for each of three stimulus sets. In the study phase, each observer saw a randomly chosen subset of 72 images. The study images were presented consecutively on the computer display, each for 3 s with no time between the images. Observers were told to memorize the images in preparation for a recognition test. The test phase followed immediately after the study phase. In the test phase, observers saw a sequence of 72 images, of which 36 were randomly chosen *old* images from the study

phase, and the remaining 36 were completely *new* images. Each test image was presented one at a time on the display until the observer responded. Observers were asked to label each image as “old” or “new” by pressing the appropriate computer key. The images remained on the screen until the response was given and feedback was provided for each test trial. All observers completed the test and study phase blocks for the three image types for a total of 216 test trials. The order of the blocks was counterbalanced across participants.

Data analysis

We report performance in terms of D' , for two reasons. First, D' is theoretically independent of an observer’s bias to respond “yes” or “no”. Second, it is normally distributed, unlike accuracy, which makes it more suitable for standard parametric statistics. We also computed c as a measure of criterion or bias. When computing D' , we added half an error to the false alarm or miss rate when those were 0 as a standard correction for cells with perfect performance (Macmillan & Creelman, 2005).

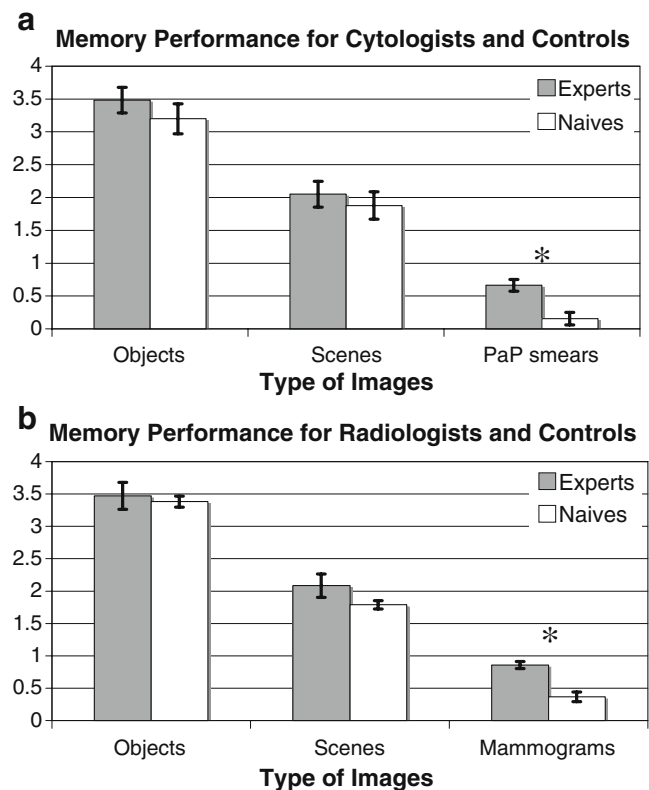


Fig. 2 Performance on visual recognition memory task by image type and group. Experts are plotted in *gray*, un-trained control observers in *white*. **a** Data from cytologists and their controls; **b** data from radiologists and their controls. *Significant difference. Error bars standard error of mean

Data were analyzed in two mixed measures ANOVAs with type of stimuli (e.g., environmental scene, object or Pap-smear/mammogram) as the within-subject factor and group membership as the between-subject factor (e.g., expert or untrained group). Planned comparisons were performed comparing the two expert and control groups on each of the three tasks using two sided *t* tests.

Results

Figures 2 and 3 show the results for two groups of experts (cytologists and radiologists) and control observers in each of the three conditions. The main finding is that neither cytologists nor radiologists were particularly good at remembering which Pap smears or mammograms they had seen ($d' < 1.0$ for both groups). Experts' memory for these images in their domain of expertise was markedly worse than their memory for objects [cytologists $t(9) = 12.43$,

$p < 0.0001$, radiologists $t(12) = 12.2$, $p < 0.0001$] or scenes [cytologists $t(9) = 8.24$, $p < 0.0001$, radiologists $t(12) = 6.21$, $p < 0.0001$]. All groups showed better memory for the object stimuli than the scene stimuli in this experiment [$F(1, 24) = 81.25$, $p < 0.0001$, $F(1,18) = 359.81$, $p < 0.0001$]. As one might imagine, experts were better than control observers with stimuli from their domain of expertise [cytologists: $t(18) = 3.88$, $p = 0.001$; radiologists: $t(24) = 4.18$, $p = 0.0003$]. Importantly, visual experts did not perform better on either scenes or objects than control observers [cytologists vs. control: $F(2, 36) = 0.70$, $p = 0.50$; radiologists vs. control: $F(2, 48) = 1.57$, $p = 0.22$]. Even if statistical power is increased by pooling data across objects and scenes and across both types of expert, performance levels of experts and controls are still statistically indistinguishable [$F(1, 45) = 1.44$, $p = 0.236$].

Are images containing an abnormality more memorable than negative cases? Positive cases were remembered better by cytology experts [$D' = 0.99$ for abnormal vs. $D' = 0.34$ for

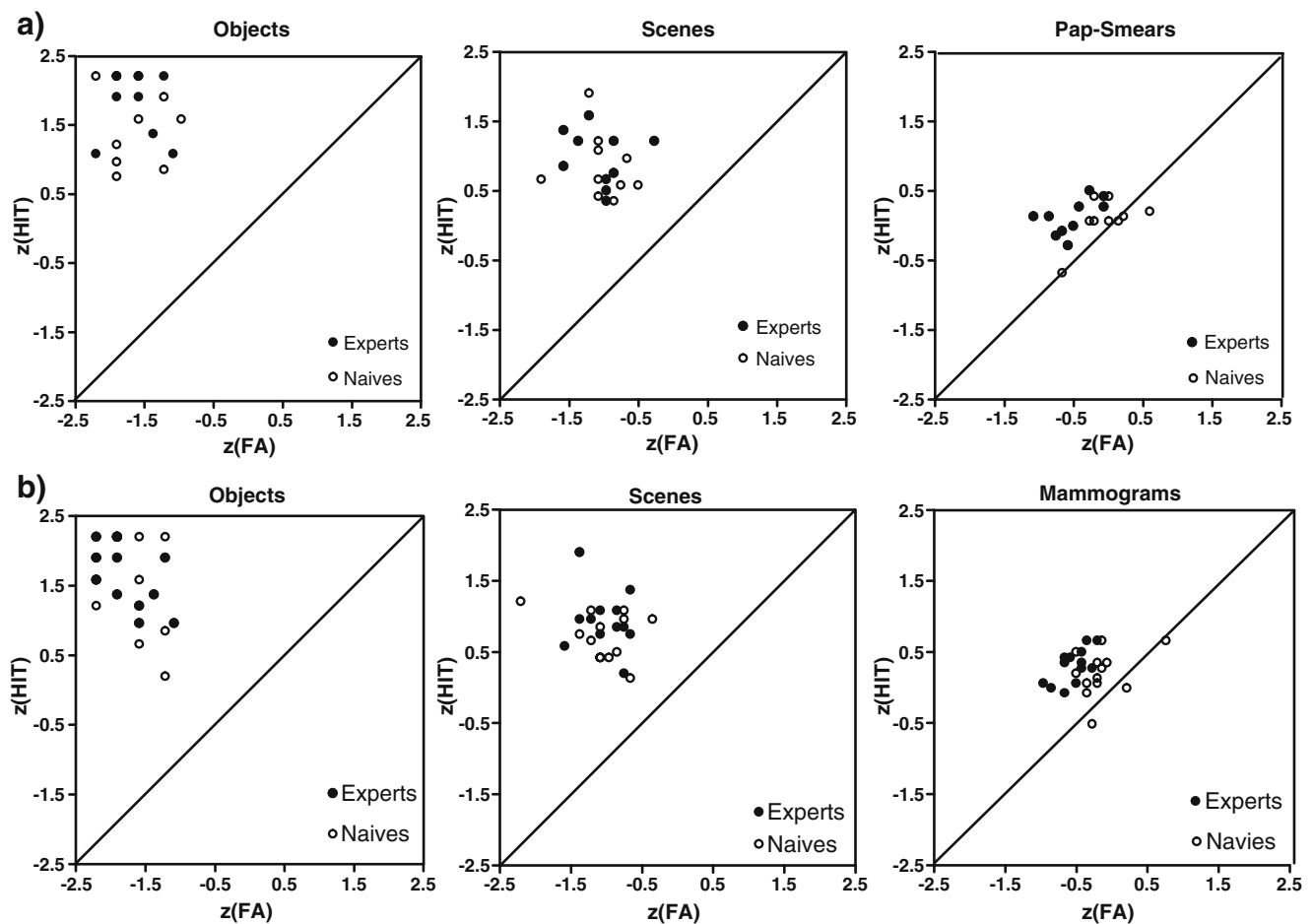


Fig. 3 Scatter plots of z score hits against z score false alarms by image type and groups. (The mean hit and false alarm rates were converted into z scores because it permits us to normalize the scores to a central mean thus allowing a comparison of measures with very

different ranges of absolute values.) Experts are plotted in *filled black circles*, un-trained control observers in *empty black circles*. **a** Data from cytologists and their controls; **b** data from radiologists and their controls

normal: $t(9) = 3.73, p = 0.005$], but not by radiologists ($D' = 0.89$ for abnormal vs. $D' = 0.88$ for normal: $t(12) = 0.08, p = 0.94$) or the control groups [control 1: $t(9) = 0.61, p = 0.554$; control 2: $t(12) = 0.86, p = 0.41$].

There were no significant differences in response criteria.

Discussion

These results show that, even though cytologists and radiologists exhibit statistically better recognition memory for images in the domain of their expertise than do medically un-trained observers, that memory is not particularly impressive and is substantially worse than their memory for natural scenes and objects. Since we do not know exactly what enables massive memory for real scenes and objects, there are several possible explanations for this finding. First, even though cytologists and radiologists might closely scrutinize up to 10,000 micrographs or mammograms a year, this experience may still be dwarfed by a lifetime of visual experience with common scenes and objects. Second, the sets of scenes and objects were more obviously heterogeneous than the medical images. Stimulus discriminability strongly affects visual recognition (Bellhouse-King & Standing, 2007; Goldstein & Chance, 1971). Even for experts, the images of cells or breast X-rays may be more similar to each other than a set objects or scenes; though recall that Brady et al. (2008) found remarkably good memory even when objects were quite similar to each other. Perhaps neither the mammograms nor the arrays of cells possess the critical attributes that would *ever* permit massive memory.

Turning to our second question, we found no significant advantage for scene or object memory in our expert populations compared to non-expert control groups. Developing a particular visual expertise did not enhance general memory for visual stimuli, nor is there evidence that individuals with particularly advanced visual memories were over-represented in our expert populations.

These data are consistent with the limited body of data on expert visual memory. Vogt and Magnussen (2005) found that artists' performance was not significantly better than that of naïves in visual recognition memory for either images of sporting events (average D' around 1.35 vs. 1.1) or images of abstract paintings (average D' around 1.2 vs. 1). However, a later study using eye tracking (Vogt & Magnussen, 2007a), found that, compared to naïves, artists showed a preference for looking at structural features as opposed to objects in the images, and they showed better memory for those features. These studies used methodologies that are not directly comparable to visual recognition memory experiments of the sort described here. Either the

images in the testing phase were very briefly presented (150 ms) and/or the set of images to be remembered was limited (20 images in the study phase). In medical imaging, one study by Hardesty et al. (2005) looked at the extent to which radiologists remember screening examinations they had interpreted before. In the study, eight radiologists were presented with 48 bilateral screening mammogram images. Of these, seven cases had been interpreted clinically by each radiologist from 14 to 38 months previously. During the first evaluation, the radiologists were given supporting documentation on the patients (e.g., prior mammograms and reports on the patient) along with mammograms they evaluated whereas in the test phase only the images without any prior reports were presented. When asked to identify the previously viewed images, none of the radiologists correctly remembered which of the mammograms they had interpreted. In the present study, experts had only the images without other supporting documentation both in the study and test phases. Our experts showed somewhat more memory than the Hardesty et al. experts, but our experts were being queried minutes, not months after the first exposure. The message from both studies is that radiologists do not have massive recognition memory for cases in the way they would have for natural scenes or objects.

In conclusion, the present data show that cytologists and radiologists were somewhat better than control observers at recognizing previously viewed stimuli from their domain of expertise. However, their absolute performance was quite poor, far worse than their performance with real scenes and objects. With those stimuli, medical experts and controls had equivalent performance. Expertise with these classes of stimuli does not allow you to remember which members of the class you have seen nor does it increase memory for visual stimuli in general.

Acknowledgments This research was funded by a NRSA F32EY019819-01 grant to Karla K. Evans and NIH-NEI EY017001 to Jeremy M. Wolfe.

References

- Allard, F., Graham, S., & Paarsalu, M. E. (1980). Perception in sport: Basketball. *Journal of Sport Psychology*, 2(1), 14–21.
- Bellhouse-King, M. W., & Standing, L. G. (2007). Recognition memory for concrete, regular abstract, and diverse abstract pictures. *Perceptual and Motor Skills*, 104(3), 758–762. doi:10.2466/PMS.104.3.758-762
- 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(38), 14325. doi:10.1073/pnas.0803390105
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436. doi:10.1163/156856897X00357
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81. doi:10.1016/0010-0285(73)90004-2

- Dallet, K., Wilcox, S., & D'Andrea, L. (1968). Picture memory experiments. *Journal of Experimental Psychology*, *76*, 312–320.
- Frey, P. W., & Adelman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, *4*(5), 541–547.
- Goldstein, A. G., & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics*, *9*(2-B), 237–241.
- Hardesty, L. A., Ganott, M. A., Hakim, C. M., Cohen, C. S., Clearfield, R. J., & Gur, D. (2005). iMemory effect in observer performance studies of mammograms. *Academic Radiology*, *12*(3), 286–290.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. NJ: Lawrence Erlbaum.
- McKeithen, K. B., Reitman, H., Judith, S., & Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers* 1. *Cognitive Psychology*, *13*(3), 307–325. doi:10.1016/0010-0285(81)90012-8
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. doi:10.1163/156856897X00366
- Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 587–595. doi:10.1037/0278-7393.15.4.587
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning & Verbal Behavior*, *6*(1), 156–163. doi:10.1016/S0022-5371(67)80067-7
- Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*(2), 207–222. doi:10.1080/14640747308400340
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74.
- Vogt, S., & Magnussen, S. (2005). Hemispheric specialization and recognition memory for abstract and realistic pictures: A comparison of painters and laymen. *Brain and Cognition*, *58*(3), 324. doi:10.1016/j.bandc.2005.03.003
- Vogt, S., & Magnussen, S. (2007a). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91. doi:10.1068/p5262
- Vogt, S., & Magnussen, S. (2007b). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, *54*(4), 298–303. doi:10.1027/1618-3169.54.4.298