Received:
22 November 2013

Revised:
31 March 2014

Accepted:
2 April 2014

doi: 10.1259/bjr.20130767

Cite this article as:

Haygood TM, Liu MAQ, Galvan E, Bassett R, Murphy Jr, WA, Ng CS, et al. Consistency of response and image recognition, pulmonary nodules. *Br J Radiol* 2014;87:20130767.

FULL PAPER

Consistency of response and image recognition, pulmonary nodules

¹T M HAYGOOD, PhD, MD, ¹M A Q LIU, ¹E GALVAN, BA, BS, ²R BASSETT, MS, ¹W A MURPHY JR, MD, ¹C S NG, MD, ¹A MATAMOROS, MD and ¹E M MAROM, MD

¹Department of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

²Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Address correspondence to: Dr Tamara Miner Haygood

E-mail: Tamara.Haygood@mdanderson.org; tamarahaygood@yahoo.com

Objective: To investigate the effect of recognition of a previously encountered radiograph on consistency of response in localized pulmonary nodules.

Methods: 13 radiologists interpreted 40 radiographs each to locate pulmonary nodules. A few days later, they again interpreted 40 radiographs. Half of the images in the second set were new. We asked the radiologists whether each image had been in the first set. We used Fisher's exact test and Kruskal-Wallis test to evaluate the correlation between recognition of an image and consistency in its interpretation. We evaluated the data using all possible recognition levels—definitely, probably or possibly included vs definitely, probably or possibly not included by collapsing the recognition levels into two and by eliminating the “possibly included” and “possibly not included” scores.

Results: With all but one of six methods of looking at the data, there was no significant correlation between consistency in interpretation and recognition of the image. When the possibly included and possibly not included scores were eliminated, there was a borderline statistical significance ($p = 0.04$) with slightly greater consistency in interpretation of recognized than that of non-recognized images.

Conclusion: We found no convincing evidence that radiologists' recognition of images in an observer performance study affects their interpretation on a second encounter.

Advances in knowledge: Conscious recognition of chest radiographs did not result in a greater degree of consistency in the tested interpretation than that in the interpretation of images that were not recognized.

In visual observer performance studies, observers' memory of images can potentially be a source of bias by preventing the observer from interpreting the image on the second or subsequent viewing(s) independently of the interpretation at the preceding viewing(s). To reduce the potential for memory-related bias, most observer performance studies allow some time to elapse between one test involving a set of images and the next. This time delay varies a great deal between studies. In a study involving the identification of rib fractures, Fuhrman et al¹ waited at least 2 years between readings. In a different study comparing detection of pulmonary nodules with varied types of monitors, Graf et al² used an average time delay of 2 days. Metz,³ advising on the design of observer performance studies, suggested that investigators separate readings by as much time as possible.

Although there is obvious concern among investigators regarding possible memory-related bias, radiologists' memory for previously encountered images has seldom been investigated. Four papers do suggest that the radiologists' ability

to recognize previously encountered experimental images may be rather limited.^{4–7}

Ultimately, the importance of memory for previously seen images in terms of design of observer studies lies not in how well radiologists consciously recognize the images that they have previously seen and distinguish them from new images but in how much their familiarity with the image affects their subsequent interpretations. This question has also seldom been addressed. Hardesty et al,⁴ studying radiologists' interpretation of mammograms that they had previously encountered clinically months earlier, found that when each radiologist's previously interpreted studies were mixed in with those of their colleagues, only one radiologist correctly recognized a single mammogram that he had previously interpreted, and then he interpreted it differently in the experiment than what he had done originally.

Kallergi et al⁸ compared interpretations by two readers of 43 combined positron-emission tomography and CT

(PET/CT) studies in patients with thyroid cancer. The readers first interpreted whole-body PET/CT. 2 days later, they interpreted the whole-body images together with dedicated images of the head and neck. At least 1 month later, they again performed the second type of interpretation with both whole-body and dedicated head-and-neck images. There were significant differences between the first and third readings and between the second and third readings but not between the first and second readings, which suggested that lingering memory of the first interpretation had impacted the second interpretation but not the third, which had occurred weeks later.

Haygood *et al*⁹ compared radiologists' consistency in interpretation of the placement of a central venous access line on repeated interpretation of images that they recognized vs those they did not recognize. There was no statistically significant difference in the likelihood of a similar interpretation on the second reading based on whether the image had been recognized. Indeed, the trend was for less consistency in interpretation for those images that were recognized than for those that were not.⁹

Our objective in this study was to examine whether conscious recognition of a chest radiograph viewed in the first half of the experiment affects the interpretation of the presence or absence and location of pulmonary nodules on a second encounter with the same chest radiograph.

METHODS AND MATERIALS

Observers and images

We showed 40 posteroanterior chest radiographs, with all patient identifying information removed, to each of the 13 radiologists who were asked to identify any pulmonary nodules and mark their location on a photocopy of the image portrayed on one-half of a standard (8.5×11 inch) sheet of paper. At the top of each image were the words "no nodules", which the radiologists circled when they thought that the radiograph contained no nodules. The images were displayed on a 3 megapixel Dome® (NDSsi, San Jose, CA) monitor using default window settings and a fit-to-screen zoom level with Stentor® 3.3 (Philips Healthcare, Andover, MA). Before doing the first reading, an appointment was made for a second reading to occur 1–3 days later. At this second reading, the same radiologists viewed 40 more frontal chest radiographs and again searched for pulmonary nodules. 20 of the 40 images in the first set had been switched out for new images, and this combination of old and new images made up the second set. The reading radiologists were told that some of the images were new and some were from the first set, but they were not told how many had been switched. The image sets were counterbalanced, so that every second reader saw the sets in the opposite order, and the individual images were randomized within the sets. At the second reading, in addition to determining the presence or absence and the location of pulmonary nodules, the radiologists were also asked whether each radiograph was new or old and how confident they were of this determination. They could grade the image as definitely, probably or possibly included in the previously viewed set vs definitely, probably or possibly not previously included.

The 60 radiographs that were included in this study were taken from 28 males and 32 females with an average age of 55.7 years. Of these, 31 radiographs contained nodules. There was 1 nodule only on 24 radiographs and 2 nodules on 7 radiographs. Nodules were confirmed by CT in all cases except one, which was confirmed by a later radiograph. They ranged in size from 7.2 to 44.2 mm (mean, 19.7 mm). Images without nodules were considered nodule free at initial interpretation and on review by two radiologists who picked the images and who did not serve as readers. They were also confirmed to be nodule free on a follow-up radiograph, both according to the initial interpretation and on review for this study. Because most of the images with nodules had incidental findings such as scoliosis or surgical clips (24 of 31), we took care that most of the nodule-free images (21 of 29) also had similar incidental abnormalities to prevent bias that might occur if only the images with nodules had such incidental findings.

The 20 repeated images included 10 with nodules and 10 without nodules. Of those with nodules, eight had a single nodule and two had two nodules. Nodules on these images ranged in size from 15.7 to 40.8 mm (average, 25.4 mm). The two radiologists who chose the images also scored the nodule-containing images based on the conspicuity of the nodules, with a score of 1 denoting a very subtle nodule, 2 denoting a nodule of intermediate subtlety and 3 denoting a relatively obvious nodule. Images with two nodules received a single score based on the more subtle of the two nodules. Of the ten repeated images with nodules, four received a score of 1, three a score of 2 and three a score of 3.

Radiologist participants were all American Board of Radiology-certified attending radiologists. They had an average of 16 years experience after residency. None were chest specialists, although all interpreted chest radiographs in the course of their practice, with the number interpreted in the previous fiscal year ranging from 29 to 9446 (average, 2103).

Scoring consistency

Nodule localizations were considered correct if the mark was placed within or touching the border of a pre-determined circle approximately 2 cm in diameter and placed such that its centre corresponded with the centre of the nodule. Any localizing mark placed outside of one of these circles or on an image without nodules was considered incorrect.

Radiologists directly indicated as described above whether they recognized each image or not. Consistency in response, however, had to be scored by the investigators. We chose two methods and used both of them. In the first method, each second interpretation was given a global score of either consistent or inconsistent. To be considered consistent, each part of the interpretation had to be identical to the interpretation rendered on the first viewing. In the second method, we gave individual scores to each part of the interpretation. If a particular nodule was identified on each interpretation, that identification would be scored as consistent. If the radiologist also marked a false nodule on one interpretation but not on the other, that mark would be graded as inconsistent, and the image as a whole would

receive the score, “consistent, inconsistent”. If the radiologist found no nodules on either viewing, the score would be “consistent” for both scoring methodologies (Figure 1).

Statistical analysis

Only the 20 repeated chest radiographs were considered. Two methods were used to classify recognition: first, the six categories were considered to be ordered, and second, the categories were collapsed into two groups: prior recognition or not. Two methods were also used to determine consistency. For the first method (see Scoring consistency section above), the Kruskal–Wallis test was used to assess the relationship between the six categories of conscious recognition of a previously interpreted image and consistency in interpretation of the image. We used Fisher’s exact test to assess the correlation between the categories collapsed into conscious recognition *vs* non-recognition of a previously seen image and consistency in its interpretation. For the second method of determining consistency, a mixed-effects logistic regression model was fitted to account for repeated measurements within patients.

Analyses were repeated excluding cases in the possibly included and possibly not included categories. We also used McNemar’s test to compare accuracy of interpretation between the first and second readings and used logistic regression to model the probability of a consistent response depending on whether nodules were present and, if they were present, their size and subtlety.

RESULTS

There were a total of 519 responses to the question about whether the images viewed during the second reading had been previously encountered (13 readers times 40 images minus one question without a recorded answer). With respect to the entire group of images in the second viewing, radiologists correctly distinguished new from old images 60.7% of the time (for additional details, please see Haygood et al¹⁰). For this project, we considered only the 260 responses for the images that were actually repeated.

Method 1 for grading consistency

There was no correlation between recognition of chest radiographs and consistency in interpretation ($p = 0.72$) (Table 1).

With the three “included” responses and the three “not-included” responses combined together, radiologists were consistent 65% of the time regarding identification of pulmonary nodules for images that they recognized, *vs* 57 % of the time for images that they did not recognize (Table 2).

If we remove the possibly included and possibly not included responses and compare consistency for the remaining, more confident, scores the proportion of consistent and inconsistent responses changes very little (Table 3).

Method 2 for grading consistency

With our second method of scoring, as each second reading of an image could receive either one or more than one score, we had 345 data points. Although the number of data points was

Figure 1. (a–f) Illustration of the scoring methods. (a, b) Show the first and second readings of an image containing one nodule. On both interpretations, the reader found and marked the nodule, and no false nodules were marked. This interpretation was scored as “consistent” by both methods used. (c, d) Show the first and second readings of an image that contains one 40.8-mm nodule in the left lower lobe, overlying the heart. In the first reading, the radiologist did not find the real nodule but marked a false nodule in the right lower lung. In the second reading, the radiologist found the real nodule and marked a different false nodule overlying the right hilum. This interpretation was scored as “inconsistent” by our first method and “inconsistent, inconsistent, inconsistent” by our second method. (e, f) Show the first and second readings of an image containing 27.5- and 26.8-mm nodules in the left upper lobe. The reading radiologist correctly identified both nodules each time, but on the second interpretation, this radiologist also marked a false nodule overlying the heart. This interpretation was scored as “inconsistent” by our first method and as “consistent, consistent, inconsistent” by our second method. In each of these cases, the reading radiologist indicated that the image was “definitely included” in the previous set of images. Arrows indicate reader marks. The black circles to the left on figures (c, d) are holes punched in the paper for storage after the data were collected.

Table 1. Response consistency regarding identification and localization of nodules for the 20 repeated images when divided by all possible levels of recognition responses

Response category	Inconsistent (n)	Consistent (n)	Total responses (n)
Definitely included	26 (35%)	49 (65%)	75
Probably included	15 (41%)	22 (59%)	37
Possibly included	11 (30%)	26 (70%)	37
Possibly not included	22 (44%)	28 (56%)	50
Probably not included	23 (43%)	31 (57%)	54
Definitely not included	3 (43%)	4 (57%)	7
Total	100 (38%)	160 (62%)	260

Our readers responded consistently 62% of the time, with a non-significant trend towards being consistent more often when they recognized the image than when they did not.

higher, the proportion of consistent and inconsistent responses was nearly the same with both scoring methods (Table 4).

Combining the definitely, probably and possibly included categories together and contrasting them with the combined definitely, probably and possibly not included categories approached but did not achieve statistical significance ($p = 0.066$). When excluding the possibly included and possibly not included categories and comparing the only remaining, more confident responses the comparison between memorable and not memorable was borderline statistically significant ($p = 0.04$), and the proportion of consistent to inconsistent responses remained about the same with 36% of pulmonary nodule identifications being inconsistent and 64% being consistent.

If we look at accuracy of interpretation for the repeated images and compare the two readings, during the first reading, our radiologists correctly identified 101 (64.7%) out of 156 nodules and correctly interpreted 86 (66.2%) out of 130 images with no nodules, our radiologists identified 144 (55.4%) out of 260

images correctly. During the second reading, our radiologists correctly identified 96 (61.5%) out of 156 nodules and correctly interpreted 100 (76.9%) out of 130 images with no nodules, they correctly interpreted 149 (57.3%) out of 260 images. There was no statistically significant difference in the accuracy of interpretation between the first and second readings ($p = 0.62$).

When evaluating factors that might influence the likelihood of a consistent response, which we based on scoring method 1, we found a significant relationship between nodule size and consistency ($p = 0.008$), with larger nodules being more likely to garner consistent interpretations. There was no statistically significant relationship between the simple presence of nodules and consistency of interpretation ($p = 0.90$), nor was there a significant relationship for nodule-containing images between the conspicuity score and consistency of interpretation ($p = 0.12$).

DISCUSSION

Most investigators allow a period of time to elapse between viewings in observer performance studies. This time lapse can differ greatly from one investigation to another and might vary from reader to reader based on such factors as equipment availability or work schedules. The time lapse is intended to avoid bias caused by image recognition, which might occur if the first interpretation colours the response on the second reading.

We have shown that in a study involving the detection and localization of pulmonary nodules on digital radiographs of the chest, in five out of six ways of looking at the data there was no statistically significant relationship between identification of the same pulmonary nodules in each of the two readings and conscious recognition of the image. We did find a trend in these five evaluations of the data towards greater consistency in nodule identification among the recognized images than among those that were not recognized, but it was not statistically significant.

Only when we excluded the “possibly included” and “possibly not included” responses and used scoring method 2 did we find a barely statistically significant result favouring greater consistency with recognized than with that of not-recognized images. We are inclined to discount the practical importance of this finding because (a) it represents only one of six ways of looking at the data; (b) it required excluding 87 (33.5%) data points using scoring method 1 and 109 (31.6%) data points using scoring method 2; and (c) the level of statistical significance was

Table 2. Response consistency regarding nodule identification and localization when sorted by combined levels of recognition

Response category	Consistency in response		
	Inconsistent (n)	Consistent (n)	Total responses (n)
Conscious recognition of repeated images			
Recognized	52 (35%)	97 (65%)	149
Not recognized	48 (43%)	63 (57%)	111
Total responses	100 (38%)	160 (62%)	260

“Recognized” includes the scores definitely, probably and possibly included.

“Not recognized” includes definitely, probably and possibly not included scores.

There is no statistically significant relationship between the two factors of recognition and consistency ($p = 0.17$).

Table 3. Response consistency regarding nodule identification and localization for the more confident levels of recognition only

Response category	Consistency in response		
	Inconsistent (<i>n</i>)	Consistent (<i>n</i>)	Total responses (<i>n</i>)
Conscious recognition of repeated images			
Definitely or probably included	41 (37%)	71 (63%)	112
Definitely or probably not included	26 (43%)	35 (57%)	61
Total responses	67 (39%)	106 (61%)	173

There is no statistically significant correlation between image recognition and consistency of interpretation ($p = 0.81$).

only $p = 0.04$. A p -value of 0.04 means that even if there is no true association, 4% of the time one would find the same amount of apparent association that we found or more.

Our findings, other than for the one borderline statistically significant result mentioned above, were similar to those in a study of central venous line position,⁹ which also disclosed no statistically significant correlation between consistency of response and conscious recognition of the image, although in that study, the trend was in the opposite direction, with a tendency for less consistency in the line position determination for recognized than that for non-recognized images. Our findings departed from the implied results of the study by Kallergi et al⁸ of PET/CT interpretation. Those investigators found greater differences in interpretation between readings done a month apart than between readings done 2 days apart. This suggested that in the second set of the readings separated by only 2 days from the first set, the interpretations were influenced by the readers' memory of their previous interpretations.⁸ That study, unlike either this one or the study of central line position,⁶ did not specifically ask readers if they remembered the images.

A few prior studies have investigated recognition of radiologic images. They found that radiologists' ability to distinguish new from repeated images on a second encounter soon after the first was just slightly better than chance.⁵⁻⁷

There are other studies, however, that have demonstrated that humans have a great capacity for remembering and distinguishing among photographs of objects. Brady et al,¹¹ found

that a few minutes after seeing photographs of 2500 objects, adults could correctly identify the previously viewed picture 87% of the time in a two-alternative forced-choice experiment, even when the new photograph differed in only a small detail (e.g. a backpack shown open *vs* closed). The two-alternative forced-choice experimental design and the widely varied types of images included in Brady's study might have contributed to the greater success of his subjects than that of ours in distinguishing the new from the old images.

For study results to be affected by recognition-related bias, the subjects must both recognize the image and remember it well enough that their second interpretation is coloured by their first interpretation. Our overall results suggest that our readers' second interpretation was not obviously coloured by their first reading, even when they consciously recognized an image. There may be several reasons for this lack of apparent influence of the first reading on the second, and different reasons could be true to greater or lesser extents for various readers and images. A reader might have recognized an image but not remembered the earlier interpretation. A reader might remember both the image and the previous interpretation but deliberately make a new decision. This may be a fairly independent decision, but it is also possible that a reader might be so intent on forming an independent interpretation on the second viewing that he/she would convince themselves of a different interpretation. We suspect this might be most likely for images that are more complex and difficult to interpret. It is also possible that the apparent influence of recognition on consistency of response was dampened by an unappreciated influence of memory on

Table 4. Response consistency regarding identification and localization of pulmonary nodules using scoring method 2

Response category	Inconsistent (<i>n</i>)	Consistent (<i>n</i>)	Total responses (<i>n</i>)
Definitely included	32 (28%)	83 (72%)	115
Probably included	17 (40%)	26 (60%)	43
Possibly included	18 (38%)	30 (63%)	48
Possibly not included	25 (41%)	36 (59%)	61
Probably not included	30 (45%)	36 (55%)	66
Definitely not included	5 (42%)	7 (58%)	12
Total responses	127 (37%)	218 (63%)	345

There is no evidence of correlation between image recognition and consistency in interpretation ($p = 0.31$).

those images that were not reported by the readers as being consciously recognized but that may have been remembered to some degree at an unconscious level.

If readers draw on their memory of prior image sets to influence their interpretation of subsequent sets, one might expect their performance on later sets to improve. This was not the case here. Specificity went up modestly, sensitivity went down slightly and overall accuracy scarcely budged.

There are other factors besides recognition of an image that might contribute to consistency of interpretation. The questions that were asked of each radiologist on viewing each image, "are there any nodules, and, if so, where are they?" had distinct correct answers. The radiologists were all well trained to answer those questions. This fact, which is also true in most observer performance studies in radiology, would tend to promote consistency regardless of recognition of an image. Larger and more obvious nodules might promote consistency simply because they would likely be noticed, and then the radiologists' training would take over. We did find this to be true regarding nodule size. We did not find it to be true regarding our opinion of the conspicuity of the nodules, which probably means at least partly that we are not as good at judging the conspicuity of a nodule as we would like to think we are. There are other factors at work as well, however, in that consistency required not merely that a radiologist find the nodule each time but also that any false-positives marked had to be identical, and that factor may have dampened any connection between the characteristics of the nodules and consistency.

On the whole, our nodules, even those that we graded a 3 for conspicuity, tended to be subtle, as one might suspect from an average sensitivity on the first and second readings of 63.9%. If we had used more obvious nodules, then consistency of interpretation would probably have been higher, but this would be expected to affect interpretation of both recognized and non-recognized images.

Our findings imply that in reader performance studies involving interpretation of single images, not only will readers remember a relatively small number of the images but this memory will also have little effect on their interpretation, even if the second interpretation occurs just a few days after the first. Although we did not specifically test clinical interpretation, we also believe that this implies that radiologists tend to bring a fresh look to each study interpreted, reading each image reasonably independently of others just interpreted or of any previously interpreted for the same patient. This is a topic that may be worth further study in the future.

One limitation to this study concerns statistical power. We studied 13 readers and their interpretations of 20 repeated images. A power calculation based on the statistical evaluation that most closely approached without reaching statistical

significance (scoring method 2 with all included and not included responses combined, $p = 0.066$) indicated that with 811 total data points (2.4 times as many as we had), we would have an 80% probability of being able to detect a 10% increase in consistency for recognized as opposed to not recognized images or vice versa. This calculation also supposed that 484 images would be in the recognized category and 327 in the not recognized category; departure from this ratio would require more total images. Achieving statistical significance for any of the other analyses would also require more images. Our design requires the readers to interpret three times as many images as are ultimately used for the consistency calculation, so at least 2433 in total. We would wish still for each reader to interpret at each reading, a number of images not unreasonable for a receiver operating characteristic study (probably no more than 60), so we would need at least 27 readers. This number of readers and observations probably exceeds practical limits.

As suggested above, we are not able to test for any recognition or influence of one reading on the other that occurs below the level of consciousness. We used two different methods of scoring consistency of interpretation. We favoured this approach because method 1 tended to emphasize inconsistency and method 2 tended to emphasize consistency, but other methods could also have been devised and might have altered the results, although we think not greatly.

Finally, we tested only one sort of image. We suspect that one reason Kallergi *et al*¹⁸ found evidence of one reading's influence over the next and we did not might be the nature of the images they showed. PET/CT examinations will consist of multiple images, and each image has the potential to harbour a finding that might trigger memory of the study as a whole. Our readers were only shown one image for each patient, and if that image did not trigger a memory, there was nothing else to do so. Therefore, our results might have been different if we had tested different types of images, particularly studies comprising a collection of multiple related images.

CONCLUSIONS

We found no convincing evidence that radiologists' recognition of a chest radiograph that they had previously interpreted coloured their interpretation in a second viewing, 1–3 days later. This suggests that they read the images on second viewing essentially independently of their first interpretation.

FUNDING

Dr Haygood was supported in part by a grant from The University of Texas–MD Anderson Cancer Center, John S. Dunn Sr, Distinguished Chair in Diagnostic Imaging. This research was also supported by the National Institutes of Health/National Cancer Institute under award number P30CA016672 and used the Biostatistics Resource Group.

REFERENCES

1. Fuhrman CR, Britton CA, Bender T, Sumkin JH, Brown ML, Holbert JM, et al. Observer performance studies: detection of single versus multiple abnormalities of the chest. *AJR Am J Roentgenol* 2002; **179**: 1551–3. doi: [10.2214/ajr.179.6.1791551](https://doi.org/10.2214/ajr.179.6.1791551)
2. Graf B, Simon U, Eickmeyer F, Fiedler V. 1K versus 2K monitor: a clinical alternative free-response receiver operating characteristic study of observer performance using pulmonary nodules. *AJR Am J Roentgenol* 2000; **174**: 1067–74.
3. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Invest Radiol* 1989; **24**: 234–45.
4. Hardesty LA, Ganott MA, Hakim CM, Cohen CS, Clearfield RJ, Gur D. “Memory effect” in observer performance studies of mammograms. *Acad Radiol* 2005; **12**: 286–90.
5. Hillard A, Myles-Worsley M, Johnston W, Baxter B. The development of radiologic schemata through training and experience. A preliminary communication. *Invest Radiol* 1985; **18**: 422–5.
6. Ryan JT, Haygood TM, Yamal JM, Evanoff M, O’Sullivan P, McEntee M, et al. The “memory effect” for repeated radiologic observations. *AJR Am J Roentgenol* 2011; **197**: W985–91.
7. Evans KK, Cohen MA, Tambouret R, Horowitz T, Kreindel E, Wolfe JM. Does visual expertise improve visual recognition memory? *Atten Percept Psychophys* 2011; **73**: 30–5. doi: [10.3758/s13414-010-0022-5](https://doi.org/10.3758/s13414-010-0022-5)
8. Kallergi M, Pianou N, Georgakopoulos A, Kafiri G, Pavlou S, Chatzioannou S. Quantitative evaluation of the memory bias effect in ROC studies with PET/CT. *Proc SPIE* 2012; **8318**: 83180D1–8.
9. Haygood TM, Ryan J, Liu QMA, Bassett R, Brennan PC. Image recognition and consistency of response. *Proc SPIE* 2012; **8318**: 83180G–1.
10. Haygood TM, Liu MAQ, Galvan EM, Bassett R, Devine C, Lano E, et al. Memory for previously viewed radiographs and the effect of prior knowledge of memory task. *Acad Radiol* 2013; **20**: 1598–603. doi: [10.1016/j.acra.2013.08.015](https://doi.org/10.1016/j.acra.2013.08.015)
11. Brady TF, Konkle T, Alvarez GA, Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci U S A* 2008; **105**: 14325–9. doi: [10.1073/pnas.0803390105](https://doi.org/10.1073/pnas.0803390105)