An expert advantage in detecting unfamiliar visual signals in noise

Zahra Hussain[a,1]

[a]Department of Psychology, American University of Beirut, Riad El-Solh, Beirut 1107 2020, Lebanon

Diagnostic radiologists are experts in discriminating and classifying medical images for clinically significant anomalies. Does their perceptual expertise confer an advantage in unfamiliar visual tasks? Here, this issue was investigated by comparing the performance of 10 radiologists and 2 groups of novices on the ability to detect novel visual signals: band-limited textures in noise. Observers performed a yes/no detection task in which texture spatial frequency and external noise levels were varied. The task was performed on two consecutive days. Contrast thresholds and response bias were measured. Contrast thresholds of radiologists were superior to the control groups in all stimulus conditions on both days. Performance improved by an equivalent amount for all groups across days. Response bias differed consistently across stimulus conditions and days but not across groups. The difference in thresholds between the radiologists and control groups suggests that experience in diagnostic medical imaging produces perceptual skills that that transfer beyond the trained domain.

radiology | domain-specific | threshold | sensitivity | bias

For the special object class of medical images, radiologists show characteristics of perceptual expertise found in other instances of highly practiced object recognition. Expert radiologists can classify medical images containing an anomaly within a fifth of a second (1–3). Radiologists make fewer and more focused eye movements than novices (4–6), localize anomalous features soon after viewing the image (5, 7), and produce fewer false positives than novices (8, 9). These perceptual feats, as in other types of naturally or professionally acquired visual expertise, have been demonstrated primarily with stimuli from the trained domain (e.g., actual or simulated X-rays or mammograms) (6, 8–11). Does the specialized perceptual experience acquired in diagnostic radiology alter how novel objects are encoded or the strategies used to detect, discriminate, or classify novel signals? Here, this question was addressed by comparing radiologists and nonradiologist control groups on the ability to detect unfamiliar visual signals in noise.

On the one hand, the typically domain-specific perceptual advantages of visual expertise should be absent for unfamiliar objects or for unpracticed tasks (12, 13). Visual experts show behavioral and neural markers of expertise only for the privileged object (birds, fingerprints) and not for objects with different spatial characteristics or configurations (13–16). Furthermore, the improvements in perceptual judgments produced with practice in laboratory-based studies of perceptual learning often are stimulus- and task-specific (17–21). From this perspective, radiologists should fare no better than novices on an unfamiliar perceptual task.

On the other hand, so-called task-irrelevant expertise emerges in conditions where stimulus or task dimensions are shared between training and transfer contexts (22). Generalized benefits of particular visual skills also have been shown when there is little in common between the trained and untrained skill, for instance, in contrast sensitivity of action videogame players (23). Additionally, under more permissive practice regimes, perceptual learning does transfer to unpracticed stimuli, tasks, and skills (20, 24–27). Therefore, the combination of visual skills acquired in medical image diagnosis may benefit certain untrained perceptual judgments.

In the experiment reported here, 10 radiologists with varying degrees of experience (Table 1) and 2 groups of novice controls detected unfamiliar signals in noise on 2 consecutive days. Control group 1 comprised a younger group of 36 individuals (primarily undergraduate students) aged between 18 and 28 y. Control group 2 comprised an older group of 10 professionals (nonradiologists) whose mean age and years of education were commensurate with the expert group. The signals were band-limited textures (Fig. 1) drawn from three spatial frequency bands, shown in two noise levels. All subjects performed the task in the six stimulus conditions.

Results

Performance. Fig. 2 shows performance of the three groups in the six conditions on both days. The top and middle rows show d' plotted against contrast for each group in each stimulus condition, averaged over subjects. The bottom row shows contrast thresholds corresponding to a d' of 1 obtained from linear fits of d' to log contrast variance. Radiologists' thresholds (shown in red) were lower than both control groups (black) on both days. Contrast thresholds were analyzed with a linear mixed-effects model in which subject was treated as a random factor, and group, spatial frequency, noise, and day were treated as fixed factors. The model included all interactions between fixed factors and first- and second-order interactions of within-subject factors spatial frequency, noise and day with subject. This formulation is equivalent to a mixed-factorial ANOVA (or a split-plot ANOVA) but uses all available data when the data are

Significance

Radiologists have years of experience in inspecting medical images for anomalies and thus are visual experts with the particular object class of medical images. Do the perceptual skills acquired in medical imaging benefit perception outside the trained domain? Here, radiologists and novice controls were compared on the ability to perform a visual detection task that was unfamiliar to all subjects. Subjects detected patterns in noise that were unlike medical images to which radiologists are routinely exposed. Radiologists were superior to the control groups in all stimulus conditions and maintained their advantage after both groups improved on the task. These results suggest that the perceptual skills developed in diagnostic radiology generalize to certain unfamiliar visual judgments.

www.manaraa.com

unbalanced and makes fewer assumptions (see *Materials and Methods* for further details). There was a significant main effect of group [$F(2, 52.98) = 4.54$; $P = 0.015$], with thresholds of the expert group lower than both control groups by a quarter to a third of a log unit (mean threshold for control 1: $8.31e^{-5}$ [95% CI: $6.76e^{-5}$ to 0.00010]; mean threshold for control 2: $9.77e^{-5}$ [95% CI: $6.6e^{-5}$ to 0.00014]; mean threshold for radiologists: $4.57e^{-5}$ [95% CI: 3.09^{-5} to $6.76e^{-5}$]). This result confirmed an effect of expertise on thresholds. There was a significant main effect of day [$F(1, 52.92) = 16.52$; $P = 0.00016$], indicating that thresholds improved across days (mean improvement: 0.2 log units; day 1 mean threshold: $8.91e^{-5}$ [95% CI: $7.24e^{-5}$ to 0.00011]; day 2 mean threshold: $5.75e^{-5}$ [95% CI: $4.67e^{-5}$ to $7.24e^{-5}$]). The effect of day was qualified by a significant day × noise interaction [$F(2, 52.11) = 5.1$; $P = 0.019$], consistent with the pattern in Fig. 2 showing greater improvement in high noise than in low noise at all spatial frequencies. A follow-up analysis on the effect of noise on the difference of log-transformed thresholds between days confirmed that thresholds improved more in high noise (mean improvement: 0.27 log units [95% CI: −0.38 to −0.16 log units]) than in low noise (mean improvement: 0.13 log units [95% CI: −0.21 to −0.06 log units]).

In addition to the effects of group and day, there was a significant main effect of spatial frequency [$F(2, 105.418) = 47.88$; $P < 0.0001$], a significant main effect of noise [$F(2, 52.273) = 927.25$; $P < 0.0001$], and a significant spatial frequency × noise interaction [$F(2, 103.80) = 13.21$; $P < 0.0001$]. The effect of spatial frequency, as expected, arose from the high spatial frequency condition at which higher contrasts were used (low spatial frequency [sf] mean threshold: $5.01e^{-5}$ [95% CI: $3.89e^{-5}$ to $6.30e^{-5}$]; medium sf mean threshold: $5.12e^{-5}$ [95% CI: $4.07e^{-5}$ to $6.60e^{-5}$]; high sf mean threshold: 0.0001 [95% CI: 0.00011 to 0.00018]). The main effect of noise, also expected, was due to higher thresholds in high noise than in low noise (low noise mean threshold: $2.18e^{-5}$ [95% CI: $1.77e^{-5}$ to $2.69e^{-5}$]; high noise mean threshold: $2.3e^{-4}$ [95% CI: 0.00019 to 0.00028]). The spatial frequency × noise interaction suggested that the difference between noise levels varied with spatial frequency. A follow-up pairwise comparison across spatial frequencies (low vs. medium, medium vs. high, low vs. high) of the difference of log-transformed low and high noise thresholds showed that the effect of noise was greater for high spatial frequencies (mean difference: 1.22 log units [95% CI: 1.02 to 1.37 log units]) than for low spatial frequencies (mean difference: 0.83 log units [95% CI: 0.69 to 0.97 log units]). In other words, the slope of threshold against noise was steeper for high spatial frequencies than for low spatial frequencies. The effect of group did not interact significantly with any variables, and none of the other interactions was significant ($F < 3$; $P > 0.05$).

Table 1. Radiologist subject information

ID	Age, y	Sex	Rank	Experience, y	Modality I	Modality II
500	45	F	Attending	16	MRI	CT
501	28	M	Resident Y5	4	CT	X-ray
502	30	M	Fellow Y6	5	CT	Ultrasound
503	29	M	Resident Y4	3	X-ray	CT
504	29	F	Resident Y4	3	CT	MRI
505	37	F	Attending	13	X-ray	Ultrasound
506	31	M	Resident Y3	2.5	CT	X-ray
507	35	F	Attending	7	CT	PET
508	30	M	Resident Y3	2	X-ray	CT
509	36	M	Attending	10	CT	PET

CT, computed tomography; F, female; M, male; MRI, magnetic resonance imaging; PET, positron emission tomography.
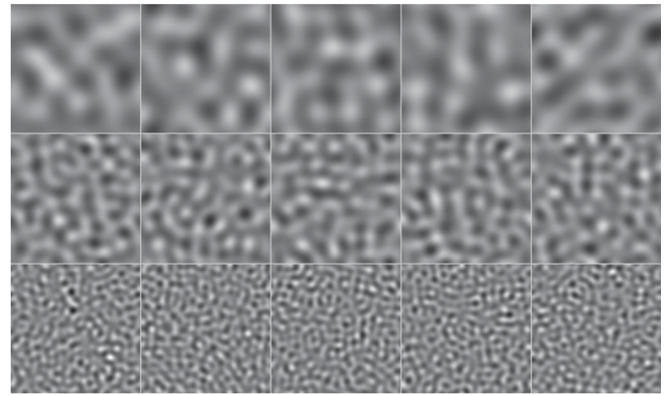

Fig. 1. Texture stimuli. Each texture was created by applying an isotropic band-pass (2 to 4, 4 to 8, and 8 to 16 cycles per image) ideal spatial frequency filter to Gaussian white noise.

Overall, these results show that radiologists' thresholds were lower than both control groups on both days and that thresholds improved significantly for all groups across days.

Response Bias. Fig. 3 shows the criterion, c, across days 1 and 2 for all stimulus conditions and groups. c was calculated as c_global, which accounts for the ensemble of signal contrasts used in each spatial frequency and noise condition (see *Materials and Methods* and refs. 28 and 29 for details).

Ten of 670 observations were removed from the data as extreme values (±3 SDs from the mean). A linear mixed-effects model with group, day, spatial frequency, and noise as fixed factors and subject as a random factor showed significant main effects of day [$F(1, 54.56) = 16.88$; $P < 0.001$], spatial frequency [$F(2, 108.76) = 8.99$; $P < 0.001$], and noise [$F(1, 53.86) = 57.72$; $P < 0.0001$]. The significant main effect of day reflected a positive criterion shift from day 1 to day 2, i.e., subjects become more conservative with practice (day 1 mean: −0.20 [95% CI: −0.26 to −0.14]; day 2 mean: −0.09 [95% CI: −0.14 to −0.03]). The significant main effect of spatial frequency arose because the criterion was more negative (more liberal) for high spatial frequencies than for low and medium spatial frequencies (low sf mean: −0.09 [95% CI: −0.15 to −0.02]; medium sf mean: −0.12 [95% CI: −0.19 to −0.06]; high sf mean: −0.21 [95% CI: −0.27 to −0.15]). The significant main effect of noise arose because the criterion was more negative in high noise than in low noise (low-noise mean: −0.01 [95% CI: −0.08 to 0.04]; high noise mean: −0.27 [95% CI: −0.34 to −0.20]). The main effect of group was not significant, and none of the interaction effects was significant ($F < 3$; $P > 0.10$).

Overall, these results indicate that subjects became more conservative with practice and that the criterion varied consistently for all groups across spatial frequencies and noise levels. Subjects were more liberal in high noise than in low noise and more liberal for high spatial frequencies than for low and medium spatial frequencies. The criterion did not differ significantly among groups.

Discussion

Radiologists were superior to novices on an unfamiliar visual-detection task performed by all subjects for the first time. The task was yes/no detection of textures of varying spatial frequency content in two levels of external noise. The controls—a younger group of undergraduate students and an older group of professional individuals—performed similarly on the task. Radiologists' thresholds were lower than both control groups, suggesting a specific effect of previous visual experience and not a general attentional or motivational advantage.
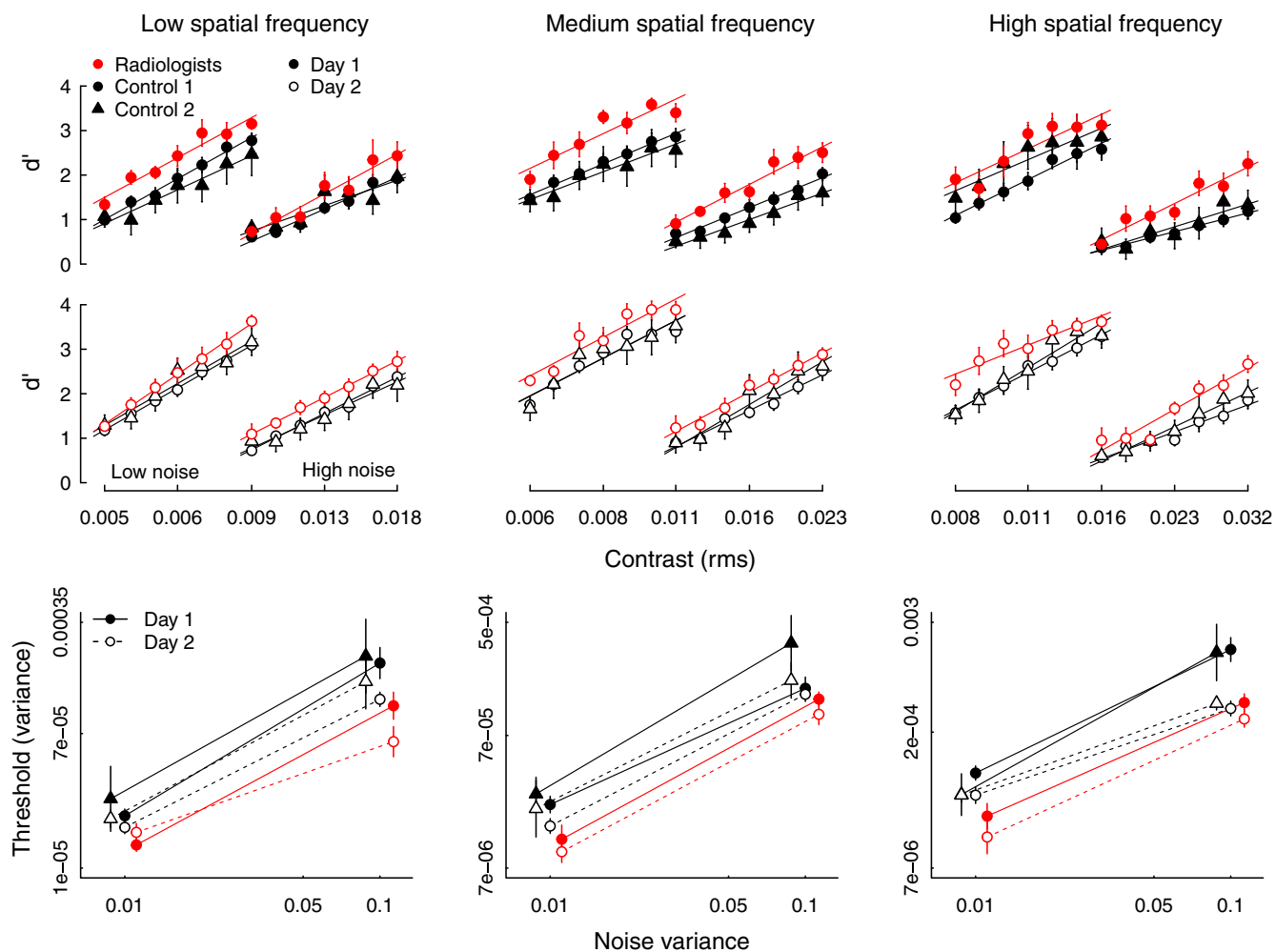
Hussain

Fig. 2. Performance of radiologists and controls on the texture detection task in each spatial frequency condition and noise level. Top and middle rows show *d′* plotted against rms contrast on day 1 and day 2. Radiologists are shown in red (*N* = 10), and the control groups are shown in black (control 1: *N* = 36; control 2: *N* = 10). The bottom row shows contrast thresholds plotted against external noise on days 1 and 2 for the two groups (*N* varies across condition; see Table 3 for details). Symbols show the mean in each condition; error bars show SEM.

All groups improved across days by an equivalent amount, confirming that the sensory skills required for the task could be acquired and were not self-selected among the experts. The stimuli and task used here were not derived from medical images or from training procedures used in diagnostic radiology. Hence, these results provide an instance of a generalized benefit of the perceptual skills acquired in diagnostic medical imaging.

No cross-domain advantage was found in previous studies where radiologists were tested on visual search for nonmedical targets or on visual memory for scenes (30, 31). Those results were interpreted as evidence against natural, preexisting differences in perceptual skills between radiologists and novices and evidence for domain-specific perceptual skills acquired in medical imaging. The radiologists tested here were not selected for training on the basis of perceptual aptitude, skill, or visual sensitivity (consistent with practice worldwide, although normal visual acuity is a criterion in some institutions). Therefore, it is assumed that all individuals in this study spanned the range of basic perceptual skill, with the group difference reflecting the generalization of learned skills acquired in professional training. The finding that practice improved performance of all groups equivalently on this task further argues against the possibility that the group difference was due to

intrinsic differences between participants. Nevertheless, longitudinal studies would be required to definitively address this question.
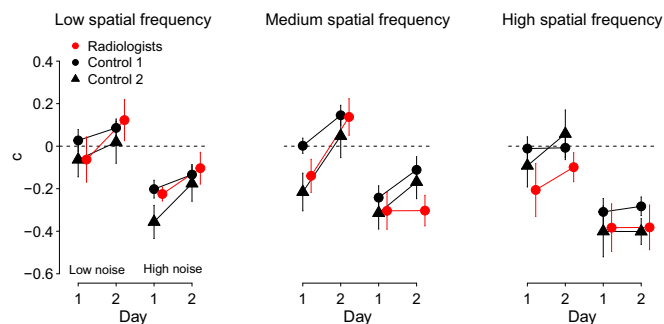

Fig. 3. Response bias of radiologists and controls in the three spatial frequency conditions (low spatial frequency [*Left*], medium spatial frequency [*Center*], and high spatial frequency [*Right*]) and two noise levels (low noise and high noise) on days 1 and 2. *c* was calculated against the ideal criterion for the ensemble of signal contrasts in each spatial frequency and noise condition (see *Materials and Methods* for details). Symbols show mean; error bars show SEM.
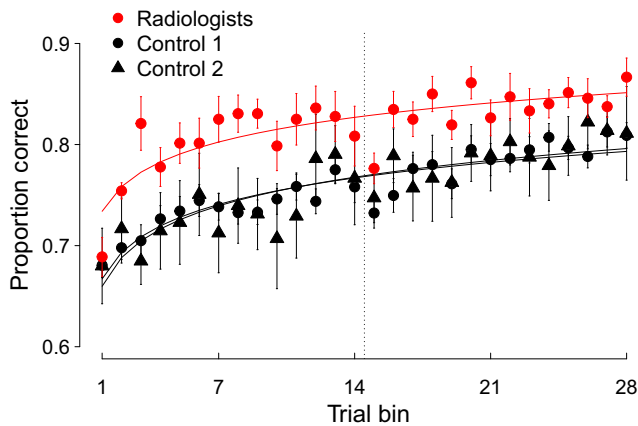
Fig. 4. Time course of performance across days 1 and 2. Proportion correct (averaged over signal-present and -absent trials) plotted for each group against trial bins comprising 12 trials each, averaged over stimulus conditions (i.e., each symbol represents 12 × 3 spatial frequency × 2 noise levels: 72 trials per subject). Solid lines show fits of proportion correct to log bin number. Vertical line separates the 2 d (14 bins per day). Symbols show mean; error bars show SEM.

The cross-domain effect of expertise shown here may derive from shared dimensions between the training and transfer contexts. Unlike the present task, diagnostic medical imaging involves the detection of spatially localized objects (e.g., lesions, microcalcifications) in structured backgrounds with non-Gaussian statistics. Signal properties and location may be known exactly or approximately, and observers utilize prior knowledge of the statistical properties of the background. With some modifications to the template (or ideal observer), performance in these tasks nevertheless may be evaluated with the same metrics and is modeled as a similar process (32–37). More work is needed to identify the parameters governing transfer of expertise beyond the trained image properties. Although one might have expected the expert effect to be larger for high spatial frequencies (finer detail), there was no significant difference in expertise across spatial frequency conditions. Visual inspection of medical images for anomalies also has been described as a two-stage process, the first comprising nonselective, Gestalt-like evaluation of the statistical properties of the image (or image gist), followed by selective, constrained search for a target (2). The visual detection task used here approximates the first stage of gist extraction, which may more readily show effects of expertise in novel contexts than selective processes such as visual search or visual memory.

Sensitivity to contrast-defined features previously has been compared between radiologists and novices. Sowden et al. (10) found that radiologists were more sensitive than novices in detecting peripherally positioned dots in X-ray images—ostensibly a familiar visual analysis performed in a novel context. Consistent with the results shown here, their novice group improved with practice on the task, suggesting that this was a learned sensory skill that could transfer from prior experience in the domain. Leong et al. (11) found no difference in contrast detection thresholds of radiologists and novices for a noiseless disk-like object located in a mammogram image, presented for unlimited duration. However, the groups differed on other performance measures such as search speed, suggesting that group differences in thresholds may have emerged under more restrictive viewing conditions (11). Compared with these previous studies, the task used here shared fewer similarities with medical images and did not rely on local feature detection. The textures were briefly presented, difficult to detect, and varied from trial to trial, preventing subjects from relying on a particular feature or on contrast differences in a particular location of

the image. Performance here relied on accessing the global spatial properties of the textures, which radiologists presumably did more swiftly than controls.

Was the expert advantage present at the very first trial, or did it emerge later in the session? Fig. 4 shows proportion correct plotted against trial bins comprising 12 trials per condition on days 1 and 2. Proportion correct is averaged over signal-present and -absent trials and over spatial frequency and noise conditions (i.e., each data point represents 72 trials per subject). The solid lines show linear fits of proportion correct to the logarithm of the bin number, which gives a reasonable approximation of the time course of improvement in perceptual tasks (38, 39). Fig. 4 suggests that all groups performed similarly in bin 1, with the group difference emerging around bins 2 to 3 and the radiologists maintaining an advantage as the session progressed. The data show a drop in performance for all groups between bins 14 and 15 (between the last bin on day 1 and the first bin on day 2), unlike the continuous improvement typically found when subjects practice with a fixed set of stimuli (19, 39). The drop between days here likely arose due to uncertainty from performing multiple stimulus conditions within the same session (i.e., subjects took a few trials to adjust to each block). A one-way ANOVA on proportion correct in bin 1 (day 1) confirmed that the groups did not differ significantly in the first bin [$F(2, 667) = 1.096; P = 0.33$]. A linear model of the full time course of learning including log bin and group as predictors showed a significant effect of group [$F(2, 1562) = 57.32; P < 0.0001$] and a significant effect of log bin [$F(1, 1562) = 211.35; P < 0.0001$]. The interaction between group and log bin was not significant [$F(2, 1562) = 0.211; P = 0.81$]. Performance in the first bin notwithstanding, radiologists' advantage emerged early within the first session and was reflected by the difference in group intercepts (control 1: 0.67; control 2: 0.66; expert: 0.73), and not slopes (slope = 0.08), as in other instances of generalized perceptual skill (40).

Practice improved performance of all groups by an equivalent amount, as reflected by the decrease in contrast thresholds across days (Fig. 2) and by the slopes of performance against bin (Fig. 4). On day 2, the control groups' thresholds were equivalent to expert thresholds on day 1. A comparison of day 1 thresholds of the expert group with day 2 thresholds of the control groups confirmed that the effect of group was not significant [$F(2, 50.691) = 1.24; P = 0.29$]. Therefore, the expertise effect on this task could be considered the equivalent of about 1,000 practice trials (approximately 170 trials per condition). The
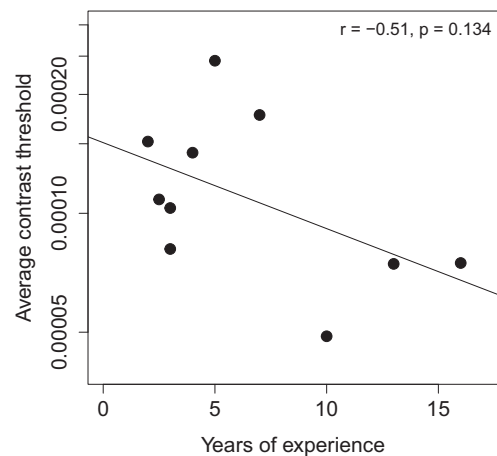

Fig. 5. Contrast thresholds (variance) of radiologists averaged over stimulus conditions and day plotted against years of professional experience. Symbols show individual subjects. Solid line shows linear fit of log threshold to years of experience. Correlation statistics are shown in the plot.

experts improved as well, however, and Fig. 4 suggests that all groups would have continued to improve past day 2. Additional practice would be needed to determine whether practice ultimately eliminates the expertise effect or whether experts and novices asymptote at different levels regardless of the amount of practice. Years of professional experience appeared to be associated with average thresholds of radiologists (Fig. 5); however, this association was not statistically significant (see *SI Appendix*, Fig. S2 for the association shown separately for each day). The relationship between years of experience and skill level, if any, is likely to be moderated by age-related declines in vision.

Did perceptual strategies differ between experts and controls? Cognitive biases relevant to diagnostic radiology have been described (41, 42), but response bias has not often been compared between experts and novices within the framework of detection theory. Although, from Fig. 3, it appears that radiologists were more liberal than controls in certain conditions, there was no significant difference in criterion between groups. Therefore, the sensitivity difference between groups was not associated with a difference in response bias. The criterion varied across stimulus conditions in the same way for experts as for novices and moved rightward (became less liberal) with practice for both groups, consistent with the effects of learning on criterion in other detection tasks (28, 29, 43). Therefore, both expertise and practice produced an increase in detection sensitivity, but only direct practice of the task altered response bias. In this respect, the cross-domain effects of visual expertise differ from direct practice or perceptual learning.

The false-positive rate often has been compared between expert radiologists and novices, with the finding that experts make fewer false positives (8, 9, 11). Here, radiologists made significantly more hits but not significantly fewer false alarms than controls (*SI Appendix*). Although this result may seem at odds with previous findings, the previous work compared experts with novices on tasks within the domain of expertise and not on unpracticed tasks. Consistent with the idea that the generalized effects of expertise may differ from the effects of direct, task-relevant practice, the practice effect on this task was driven by a change in false alarms across days more so than an increase in hits (*SI Appendix*, Fig. S1).

Overall, the results of this study support a generalized advantage of visual expertise arising from professional training in diagnostic radiology. Here, this advantage was shown for visual detection in noise, which resembles the perceptual tasks in medical imaging but which nevertheless was novel in stimulus type and task procedure. The scope of expertise is likely to depend on the exact perceptual skills trained and the properties of the task on which generalization is tested. Would expertise in a different visual domain, such as fingerprint-matching, produce a benefit for the sort of task used here? Medical imaging and fingerprint-matching undoubtedly involve distinct, domain-specific skills, but perhaps common perceptual judgments are honed across different varieties of visual expertise.

Materials and Methods

This study was approved by the Institutional Review Board of the American University of Beirut (AUB). The procedure was explained to subjects in advance, and all subjects gave informed consent before participating.

Subjects. The radiologists were attending physicians or interns with a minimum of 2 y of training in the Department of Diagnostic Radiology at the AUB Medical Center. Table 1 gives the age, sex, years of experience, and primary imaging modalities of the radiologist subjects. The control groups comprised primarily students (control 1) and faculty members at AUB (control 2). Table 2 gives the ages of the control groups. The mean age of the radiologists was 33 y (SD = 5.3 y), control 1 was 20 y (SD = 2.5), and control group 2 was 35 y (SD = 3.3). Four of 10 radiologists, 19 of 36 subjects from control 1, and 7 of 10 subjects from control 2 were female. Eight of 10 subjects from control 2 had PhD degrees, and two had Master's degrees.

All subjects had normal or corrected-to-normal visual acuity as measured by the Early Treatment Diabetic Retinopathy Study (ETDRS) acuity chart. Eleven radiologists and 53 controls participated in the study. Four control subjects did not complete the experiment (did not return on day 2). Three controls and one radiologist misunderstood the instructions or could not perform the task (i.e., d' was 0 or less in all stimulus conditions). These subjects were excluded from the study. Hence, the final group of subjects comprised 10 radiologists, 36 subjects in control 1, and 10 subjects in control 2.

Apparatus and Stimuli. Stimuli were generated on an Intel Skull Canyon NUC computer using MATLAB Version 9.4 (MathWorks) and the Psychophysics and Video Toolboxes (44, 45). Stimuli were displayed on cathode-ray tube monitors (M783P and M783S; Dell) set to a resolution of 1,024 × 768 pixels and a frame rate of 60 Hz (noninterlaced).

The textures were band-limited noise patterns created by applying an isotropic band-pass ideal spatial frequency filter to Gaussian noise (1). Three spatial frequency bands were used (2 to 4, 4 to 8, and 8 to 16 cycles per image), and five textures were created for each spatial frequency condition. The textures subtended 4.8 × 4.8 degree of visual angle from the viewing distance of 114 cm. The stimuli were shown in one of two levels of static, white, two-dimensional Gaussian noise created by sampling from distributions with contrast variances of 0.01 and 0.1. During the experiment, stimulus contrast in each noise condition was varied across trials using the method of constant stimuli. For the signal-present condition, the stimulus was shown at one of seven levels of contrast spaced equally on a logarithmic scale. For the signal-absent condition (50% of trials), signal contrast was set to zero. Hence, there were eight contrast levels in all, and 48 different stimulus conditions (three spatial frequencies × two external noise levels × eight contrasts).

Procedure. All subjects performed two sessions of the yes/no detection task at roughly the same time on two consecutive days. Viewing was binocular, and head position was stabilized with an adjustable chin rest. The stimulus display was the only source of illumination in the room. The experiment started after a 60-s period, during which the subject adapted to the average luminance of the display. Each trial began with the presentation of a black, high-contrast fixation point (0.15 × 0.15 degree of visual angle), in the center of the screen for 100 ms. This was followed by a randomly selected texture (from the set of five textures) presented for 200 ms at the center of the screen in either the signal-absent condition (zero contrast) or the signal-present condition (one of seven contrasts) at the given noise level. On signal-absent trials, the stimulus comprised a square patch of Gaussian noise. After the stimulus disappeared, subjects used a keypress to report whether the texture was present or absent on that trial. Auditory feedback indicated whether the response was correct (high-pitched tone) or incorrect (low-pitched tone). The next trial began 1 s after the presentation of feedback.

Spatial frequency conditions were blocked, and noise levels were blocked within each spatial frequency condition (3 × 2 = 6 blocks). The order of spatial frequency and noise blocks was randomized for each subject. Each block comprised 168 trials, with 84 signal-absent trials and 84 signal-present trials (7 contrasts × 12 trials per contrast). Hence, the session comprised a total of 1,008 trials (168 trials × 6 blocks), over a duration of about 60 min. There was a short break between spatial frequency blocks.

Table 2. Number of subjects per age (years) for each group

Control 1		Control 2		Experts	
Age, y	N	Age, y	N	Age, y	N
18	17	30	1	28	1
19	4	33	2	29	2
20	5	34	1	30	2
21	1	35	4	31	1
22	1	39	1	35	1
23	1	42	1	36	1
24	3			37	1
25	2			45	1
26	1				
28	1				
20 (2.5)		35 (3.3)		33 (5.3)	

Mean ages in years (SD) are shown in the bottom row.

Table 3. Number of subjects per cell for threshold analyses

	Low (2 to 4 cpi)		Medium (4 to 8 cpi)		High (8 to 16 cpi)	
	Low	High	Low	High	Low	High
Radiologists, day 1	10	9	9	10	10	10
Radiologists, day 2	10	10	10	10	8	10
Control 1, day 1	36	36	33	35	35	33
Control 1, day 2	36	36	34	36	30	31
Control 2, day 1	9	10	9	9	9	10
Control 2, day 2	9	10	10	10	9	9

cpi, cycles per image.

Dependent Measures. The proportion of hits (H) was calculated at every contrast in every spatial frequency and noise condition, i.e., seven hits rates per block. The proportion of false alarms (FA) was calculated from the signal-absent trials in every spatial frequency and noise block, i.e., 1 FA per block. H and FA were used to calculate signal detection measures of d' and bias.

The signal detection measure of sensitivity, d', was calculated at each contrast using the standard formula

$$d' = z(H) - z(FA),\qquad [1]$$

where z is the inverse of the cumulative normal distribution function.

Contrast detection thresholds corresponding to a d' of 1 were obtained from a linear fit of d' to log contrast variance.

The response criterion, here termed c, was computed as c_global (28), a useful measure of criterion where multiple signal levels are interleaved with signal-absent trials in method of constant stimuli:

$$c = \lambda\_obs - \lambda\_ideal \qquad [2]$$

or

$$c = -z(FA) - \arg\max_{\lambda}\left(\sum_{i=1}^{m}(P(S_i)[1 - \Phi(\lambda; d'_i, 1)]) + P(N)[\Phi(\lambda; 0, 1)]\right),\qquad [3]$$

where $P(S_i)$ is the probability of the ith signal (i.e., 0.5/7), $P(N)$ is the probability of a noise trial (i.e., 0.5), and Φ is the cumulative Gaussian thresholded at the ideal criterion, λ, and assuming equal, unit variance for both distributions. The ideal criterion is assumed to maximize proportion correct for the ensemble of stimulus contrasts. Eq. **3** finds λ_{ideal}, and estimates the observer's criterion, λ_{obs}, as the distance from λ_{ideal} using the observer's false-alarm rate. This formulation is an extension of the standard formula for criterion:

$$c = -0.5 * z(H) + z(FA).\qquad [4]$$

Whereas in Eq. **4**, the ideal criterion maximizes proportion correct at a single signal level, the ideal criterion in Eq. **3** accounts for all signal levels.

Statistical Analyses. For a subset of subjects in each condition, reliable thresholds could not be obtained (i.e., $threshold > 0.1$ or $< 1e-6$), and data for one subject in the high spatial frequency condition on day were lost due to electrical failure. No more than two subjects were removed per cell except for the high spatial frequency conditions on day 2 in which a larger number of control 1 subjects were excluded. Table 3 shows the number of subjects per cell for whom reliable thresholds were obtained, corresponding to the threshold plots in Fig. 2.

Log-transformed thresholds were analyzed with a linear mixed-effects model in which subject was treated as a random factor, and group, spatial frequency, noise, and day were treated as fixed factors. The model included all interactions between fixed factors and first- and second-order interactions of within-subject factors spatial frequency, noise, and day with subject. This formulation is equivalent to a mixed-factorial ANOVA (or a split-plot ANOVA) and uses all available data when the data are unbalanced (46). Significant main effects were followed by Tukey pairwise comparisons, and significant interaction effects were followed up with tests of simple effects and post hoc comparisons.

Data Availability. The data reported in this paper are available upon request from Z.H.

1. H. L. Kundel, C. F. Nodine, Interpreting chest radiographs without visual search. *Radiology* **116**, 527–32 (1975).
2. T. Drew, K. Evans, M. L. H. Võ, F. L. Jacobson, J. M. Wolfe, Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics* **33**, 263–274 (2013).
3. H. Sheridan, E. M. Reingold, The holistic processing account of visual expertise in medical image perception: A review. *Front. Psychol.* **8**, 1620 (2017).
4. H. L. Kundel, P. S. La Follette, Jr, Visual search patterns and experience with radiological images. *Radiology* **103**, 523–528 (1972).
5. E. A. Krupinski, A. R. Graham, R. S. Weinstein, Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Hum. Pathol.* **44**, 357–364 (2013).
6. E. M. Kok, "Developing visual expertise: From shades of grey to diagnostic reasoning in radiology," PhD thesis, Maastrich University, Maastricht, The Netherlands (2016).
7. D. P. Carmody, C. F. Nodine, H. L. Kundel, An analysis of perceptual and cognitive factors in radiographic interpretation. *Perception* **9**, 339–344 (1980).
8. E. E. Christensen *et al.*, The effect of search time on perception. *Radiology* **138**, 361–365 (1981).
9. C. F. Nodine *et al.*, How experience and training influence mammography expertise. *Acad. Radiol.* **6**, 575–585 (1999).
10. P. T. Sowden, I. R. Davies, P. Roling, Perceptual learning of the detection of features in x-ray images: A functional role for improvements in adults' visual sensitivity? *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 379–390 (2000).
11. D. L. Leong *et al.*, Radiologist experience effects on contrast detection. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **31**, 2328–2333 (2014).
12. K. M. Curby, I. Gauthier, To the trained eye: Perceptual expertise alters visual processing. *Top Cogn. Sci.* **2**, 189–201 (2010).
13. C. M. Bukach, W. S. Phillips, I. Gauthier, Limits of generalization between categories and implications for theories of category specificity. *Atten. Percept. Psychophys.* **72**, 1865–1874 (2010).
14. O. E. Krigolson, L. J. Pierce, C. B. Holroyd, J. W. Tanaka, Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *J. Cognit. Neurosci.* **21**, 1834–1841 (2009).
15. S. Hagen, Q. C. Vuong, L. S. Scott, T. Curran, J. W. Tanaka, The role of spatial frequency in expert object recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **42**, 413–422 (2016).
16. L. Rourke, L. C. Cruikshank, L. Shapke, A. Singhal, A neural marker of medical visual expertise: Implications for training. *Adv. Health Sci. Educ. Theory Pract.* **21**, 953–966 (2016).
17. A. Fiorentini, N. Berardi, Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vis. Res.* **21**, 1149–1158 (1981).
18. R. E. Crist, M. K. Kapadia, G. Westheimer, C. D. Gilbert, Perceptual learning of spatial localization: Specificity for orientation, position, and context. *J. Neurophysiol.* **78**, 2889–2894 (1997).
19. Z. Hussain, A. B. Sekuler, P. J. Bennett, Contrast-reversal abolishes perceptual learning. *J. Vis.* **9**, 20.1–20.8 (2009).
20. Z. Hussain, P. V. McGraw, A. B. Sekuler, P. J. Bennett, The rapid emergence of stimulus specific perceptual learning. *Front. Psychol.* **3**, 226 (2012).
21. T. Saffell, N. Matthews, Task-specific perceptual learning on speed and direction discrimination. *Vis. Res.* **43**, 1365–1374 (2003).
22. Y. K. Wong, J. R. Folstein, I. Gauthier, Task-irrelevant perceptual expertise. *J. Vis.* **11**, 3 (2011).
23. R. Li, U. Polat, W. Makous, D. Bavelier, Enhancing the contrast sensitivity function through action video game training. *Nat. Neurosci.* **12**, 549–551 (2009).
24. L. Q. Xiao *et al.*, Complete transfer of perceptual learning across retinal locations enabled by double training. *Curr. Biol.* **18**, 1922–1926 (2008).
25. T. Mastropasqua, J. Galliussi, D. Pascucci, M. Turatto, Location transfer of perceptual learning: Passive stimulation and double training. *Vis. Res.* **108**, 93–102 (2015).
26. Y. Zhang, Y. F. Yuan, X. He, G. L. Zhang, Bottom-up and top-down factors of motion direction learning transfer. *Conscious. Cognit.* **74**, 102780 (2019).
27. J. Deveau, D. J. Ozer, A. R. Seitz, Improved vision and on-field performance in baseball through perceptual learning. *Curr. Biol.* **24**, R146–R147 (2014).
28. P. R. Jones, D. R. Moore, D. E. Shub, S. Amitay, The role of response bias in perceptual learning. *J. Exp. Psychol. Learn. Mem. Cogn.* **41**, 1456–70 (2015).
29. Z. Hussain, P. J. Bennett, Perceptual learning of detection of textures in noise. *J. Vis.* **20**, (2020).
30. C. F. Nodine, E. A. Krupinski, Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Acad. Radiol.* **5**, 603–612 (1998).
31. K. K. Evans *et al.*, Does visual expertise improve visual recognition memory? *Atten. Percept. Psychophys.* **73**, 30–35 (2011).
32. M. P. Eckstein, J. S. Whiting, Visual signal detection in structured backgrounds. i. effect of number of possible spatial locations and signal contrast. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **13**, 1777–1787 (1996).

Hussain

www.manaraa.com

33. M. P. Eckstein, A. J. Ahumada, Jr, A. B. Watson, Visual signal detection in structured backgrounds. ii. effects of contrast gain control, background variations, and white noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **14**, 2406–2419 (1997).

34. A. E. Burgess, F. L. Jacobson, P. F. Judy, Human observer detection experiments with mammograms and power-law noise. *Med. Phys.* **28**, 419–437 (2001).

35. Y. Zhang, C. K. Abbey, M. P. Eckstein, Adaptive detection mechanisms in globally statistically nonstationary-oriented noise. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **23**, 1549–1558 (2006).

36. C. Castella *et al.*, Mass detection on mammograms: Influence of signal shape uncertainty on human and model observers. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **26**, 425–436 (2009).

37. A. E. Burgess, Visual perception studies and observer models in medical imaging. *Semin. Nucl. Med.* **41**, 419–436 (2011).

38. B. A. Dosher, Z. L. Lu, The functional form of performance improvements in perceptual learning: Learning rates and transfer. *Psychol. Sci.* **18**, 531–539 (2007).

39. Z. Hussain, A. B. Sekuler, P. J. Bennett, How much practice is needed to produce perceptual learning? *Vis. Res.* **49**, 2624–2634 (2009).

40. Z. Liu, D. Weinshall, Mechanisms of generalization in perceptual learning. *Vis. Res.* **40**, 97–109 (2000).

41. C. S. Lee, P. G. Nagy, S. J. Weaver, D. E. Newman-Toker, Cognitive and system factors contributing to diagnostic errors in radiology. *AJR Am. J. Roentgenol.* **201**, 611–617 (2013).

42. L. P. Busby, J. L. Courtier, C. M. Glastonbury, Bias in radiology: The how and why of misses and misinterpretations. *Radiographics* **38**, 236–247 (2018).

43. B. Xu, L. Rourke, J. K. Robinson, J. W. Tanaka, Training melanoma detection in photographs using the perceptual expertise training approach. *Appl. Cognit. Psychol.* **30**, 750–756 (2016).

44. D. H. Brainard, The psychophysics toolbox. *Spatial Vis.* **10**, 433–436 (1997).

45. D. G. Pelli, The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vis.* **10**, 437–442 (1997).

46. S. E. Maxwell, H. D. Delaney, K. Kelley, *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Routledge, New York, ed. 3, 2018).

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Hussain

www.manaraa.com