

Looking in the Same Manner but Seeing it Differently: Bottom-up and Expertise Effects in Radiology

Citation for published version (APA):

Kok, E. M., De Bruin, A. B. H., Robben, S. G. F., & Van Merriënboer, J. J. G. (2012). Looking in the Same Manner but Seeing it Differently: Bottom-up and Expertise Effects in Radiology. *Applied Cognitive Psychology*, 26(6), 854-862. <https://doi.org/10.1002/acp.2886>

Document status and date:

Published: 01/01/2012

DOI:

[10.1002/acp.2886](https://doi.org/10.1002/acp.2886)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Looking in the Same Manner but Seeing it Differently: Bottom-up and Expertise Effects in Radiology

ELLEN M. KOK^{1*}, ANIQUE B. H. DE BRUIN¹, SIMON G. F. ROBBEN² and JEROEN J. G. VAN MERRIËNBOER¹

¹School of Health Professions Education, Department of Educational Research and Development, Maastricht University, Maastricht, The Netherlands

²Department of Radiology, Maastricht University Medical Center, Maastricht, The Netherlands

Summary: Models of expertise differences in radiology often do not take into account visual differences between diseases. This study investigates the bottom-up effects of three types of images on viewing patterns of students, residents and radiologists: focal diseases (localized abnormality), diffuse diseases (distributed abnormality) and images showing no abnormalities (normal). Participants inspected conventional chest radiographs while their eye movements were recorded. Regardless of expertise, in focal diseases, participants fixated relatively long at specific locations, whereas in diffuse diseases, fixations were more dispersed and shorter. Moreover, for students, dispersion of fixations was higher on diffuse compared with normal images, whereas for residents and radiologists, dispersion was highest on normal images. Despite this difference, students showed relatively high performance on normal images but low performance on focal and diffuse images. Viewing patterns were strongly influenced by bottom-up stimulus effects. Although viewing behavior of students was similar to that of radiologists, they lack knowledge that helps them diagnose the disease correctly. Copyright © 2012 John Wiley & Sons, Ltd.

INTRODUCTION

As in many other domains of visual expertise, such as meteorology (Canham & Hegarty, 2010), biological classification (Jarodzka, Scheiter, Gerjets, & Van Gog, 2010) and aviation (Remington, Johnston, Ruthruff, Gold, & Romera, 2000), diagnosing medical images requires an intricate interplay between cognitive and perceptual processes (Krupinski, 2010). Experienced radiologists have extensive knowledge of clinically normal exemplars as well as abnormal features that signal pathology (Norman, Coblenz, Brooks, & Babcock, 1992). This enables a radiologist to decide on, for example, the identity of a white area on a chest radiograph: Is this a tumor, pneumonia, or is it just an artifact of the way the image is produced? In order to gather all this information on which a diagnostic decision can be based, visual search has to take place.

Viewing behavior of experts and novices in radiology has been extensively studied (for a recent review, see Nodine & Mello-Thoms, 2010). A lot of research is devoted to the effect of image features on viewing behavior. For example, effects on viewing have been investigated for lesion conspicuity (Krupinski, 2005; Leong, Nicolaou, Emery, Darzi, & Yang, 2007; Manning, Ethell, & Donovan, 2004; Mello-Thoms, et al., 2005), nodule size (Krupinski, Berger, Dallas, & Roehrig, 2003), image quality (Krupinski & Roehrig, 2010) and local anatomical variation (Samei, Flynn, Peterson, & Eyler, 2003). However, most *expertise* research in radiology does not take into account visual variations within a stimulus-type and is conducted within the context of one specific type of disease, such as chest nodules (Manning, Ethell, Donovan, & Crawford, 2006) or tumors in mammograms (Kundel, Nodine, Conant, & Weinstein, 2007). Yet, within

one imaging modality, several types of diseases can be present, which could strongly influence viewing behavior in a bottom-up fashion (Kok, De Bruin, Robben, & Van Merriënboer, in press). The current study adds to the expertise literature in radiology (and possibly other domains of visual expertise) by investigating how those bottom-up effects of type of disease influence eye-movement patterns in novices, intermediates and experts.

Viewing behavior can be investigated using eye tracking. The most important eye movements are fixations and saccades. During a fixation, the eye remains still and takes in information. Saccades are jumps between fixations, during which no information intake occurs (Rayner, 1998). It is known that eye movements, which reflect deployment of attention, can be influenced by *bottom-up* effects and *top-down* effects. When stimulus characteristics influence eye movements, this is called bottom-up processing (Itti & Koch, 2001); when cognitive relevance strongly guides visual search, this is called top-down processing (Yarbus, 1967).

Top-down effects on viewing can arise from expectations, the specific task at hand, and also from expertise. The characteristics of expertise have been extensively studied (Ericsson, Charness, Feltovich, & Hoffman, 2006). This line of research started with the early work of De Groot, who investigated expertise in chess (de Groot, 1946). Links were made between expertise in chess (Chase & Simon, 1973) and the medical field; and shortly after, the nature of medical expertise also became the focus of expertise research (Groen & Patel, 1988; Lesgold et al., 1988; Norman et al., 1992; Schmidt, Norman, & Boshuizen, 1990). Experts possess complex cognitive structures in which knowledge is stored and organized (van Merriënboer & Sweller, 2005). In the context of medical expertise, clinical and diagnostic information is often structured in illness scripts (Schmidt & Boshuizen, 1993). Illness scripts hold extensive information related to a disease or a class of diseases, such as signs and symptoms, consequences and context under which the illness develops. Development of illness scripts is believed to

*Correspondence to: Ellen M. Kok, School of Health Professions Education, Department of Educational Development and Research, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands.
E-mail: e.kok@maastrichtuniversity.nl

occur through the process of chunking, which takes place through repeated exposure to the same symptoms with a specific diagnosis. This eventually leads to direct activation of the diagnosis when a patient exhibits those symptoms (Ericsson & Kintsch, 1995). Although working memory is severely limited in the amount of information elements that it can hold and manipulate, scripts can be held in working memory as one element. This allows experts to circumvent working memory constraints that novices experience when trying to hold a lot of information in mind (van Gog, Ericsson, Rikers, & Paas, 2005). Not only does the availability of illness scripts lower the cognitive load imposed by the visual diagnostic task, it also influences visual search in a top-down manner: experts' knowledge influences their viewing behavior.

Gegenfurtner, Lehtinen and Saljö (2011) reviewed three more specific theories of visual expertise in terms of eye-movements: The holistic model of image perception (Kundel et al., 2007), the theory of long-term working memory (Ericsson & Kintsch, 1995) and the information reduction theory (Haider & Frensch, 1999). Kundel and Nodine (1983) were among the first to investigate the perceptual aspects of visual expertise in radiology. Their holistic model of image perception describes how the expert's extensive knowledge base informs his/her viewing behavior. Upon seeing a medical image, experts gather an initial impression of the image (Kundel et al., 2007) and detect which areas are perturbed (some radiologists report that an abnormal area seems to leap out of an image); those areas are scrutinized for features of pathology. The main advantage that experts have over novices is that their initial impression is more informative. This leads to fixation of perturbations within the first 2 seconds of viewing and to longer saccades (Gegenfurtner et al., 2011).

Ericsson and Kintsch (1995) pose that expertise leads to a change in memory structures, leading to 'long-term working memory'. Information related to the field of expertise is stored in a structured manner, and is easily available for retrieval. This allows experts to encode and retrieve task-relevant information more quickly compared with novices, resulting in shorter fixation durations. The information-reduction hypothesis (Haider & Frensch, 1999) poses that experts performing a task ignore task-irrelevant information at perceptual level. That is, when approaching a visual task, such as diagnosing a chest radiograph, an expert has the ability to distinguish task-relevant information (e.g. the abnormality) from task-irrelevant information (e.g. the healthy lung tissue) and redundant information (e.g. the liver is usually redundant in chest radiographs), and to limit visual processing to relevant information only. Eye-movements reflect this selective processing: Experts are predicted to have more fixations on task-relevant information, and these fixations are of longer duration. On task irrelevant and redundant information, they show less fixations, and these are of shorter duration. Thus, experts take in mainly relevant information and less irrelevant information. This lowers the amount of information that is to be processed in working memory, thus lowering the cognitive load for experts. Novices are not yet able to distinguish task-relevant from task-irrelevant information, their working memory is further taxed by irrelevant and redundant information that they also take in, leading to higher cognitive load. For example, a lot of novices pay

attention to the air bubble that is present in the stomach of many healthy patients. This taxes their working memory, leaving less space to process disease-relevant information in working memory. The information-reduction hypothesis thus predicts that experts fixate relatively more and longer on relevant information and fixate relatively little and for a shorter time on irrelevant information.

Measuring expertise-effects on eye-movements

The three theories of visual expertise mentioned earlier each predict eye-movement differences between experts, intermediates and novices. However, measuring the time to first fixating relevant information (predicted by the holistic model of image perception) and the proportion of time on relevant information (predicted by the information-reduction hypothesis) requires that this information is localized. Part of the information that is present has to be irrelevant or redundant, otherwise the measures do not make sense. But this is not always the case. Relevant information can also be present on a more global level (Kok et al., in press; O'Neill et al., 2011; Vogt & Magnussen, 2007). When focusing on chest radiographs (X-ray images), a distinction can be made between focal diseases and diffuse diseases (Kok et al., in press), see Figure 1 for examples. Focal diseases consist of an abnormality at a specific location, such as a tumor. The rest of the lung is relatively unaffected. For example, the focal disease in Figure 1 shows a large, round abnormality in the middle of the right lung (left side of the image). The rest of the lung appears normal. In contrast to focal diseases, diffuse diseases involve the whole lung. Figure 1 shows an example of a diffuse disease, in which both lungs appear spotted and slightly whiter than normal. Images without abnormalities are usually referred to as 'normal images'. In radiological reports, focal diseases are typically described as objects (e.g. a space-occupying process), so location, size, form, and so forth are noted. In contrast, diffuse diseases are described as pattern, for instance, a reticular pattern (innumerable lines that together resemble a net) or a honeycombing pattern (a pattern of small rings, resembling a honeycomb) (Hansell et al., 2008).

These stimulus-level differences between diseases influence eye movements in a bottom-up fashion. The information that is necessary for the diagnosis is differently distributed over the image in the two types of diseases, which might lead to differences in viewing behavior between the two types of diseases and the normal images. Furthermore, the eye-movement measures of expertise that are described earlier are not all applicable to globally present information. Although the average fixation duration and saccadic amplitude could be investigated for diffuse diseases, it does not make sense to measure time to the first fixation of a diffuse disease, or the relative time on relevant information for a diffuse disease, as this disease is globally present. A different measure is necessary to investigate expertise effects for globally present relevant information. A similar issue was experienced by Vogt and Magnussen (2007), who investigated eye movement patterns of artists and laypeople on different types of images. They calculated a global/local ratio based on the distance between fixations. A higher ratio indicates a global, dispersed viewing pattern

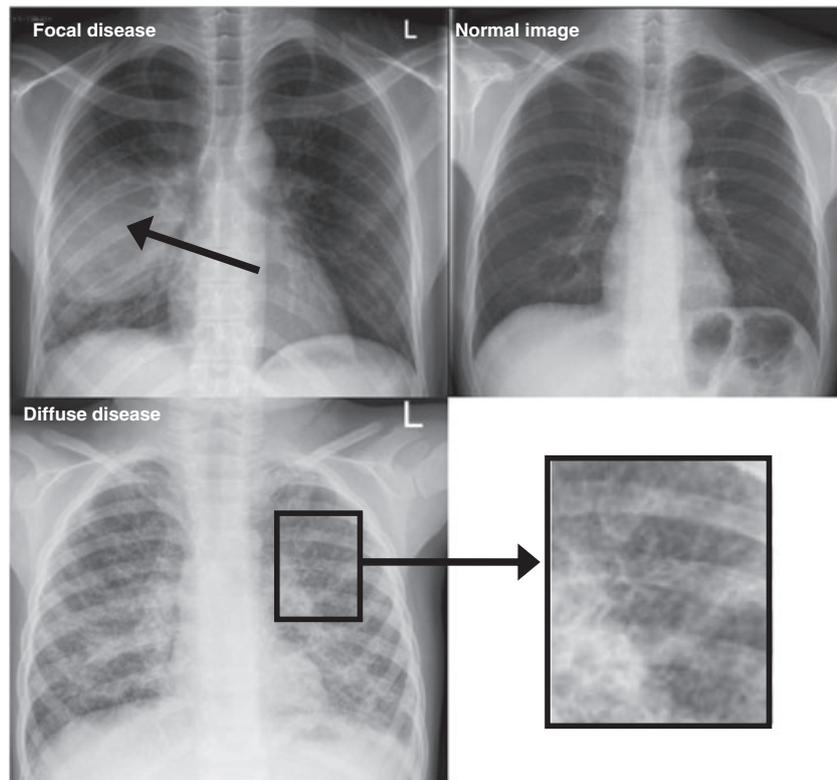


Figure 1. Example of images showing a focal disease, a diffuse disease and a normal image. Images are edited to enhance apprehensibility

whereas a lower ratio indicates that fixations cluster locally, in specific (informative) regions (Zangemeister, Sherman, & Stark, 1995). This study showed that in the field of art, both bottom-up effects of the type of image (abstract patterns vs realistic objects) and expertise influence eye movement patterns. Both artists and laypeople showed a more global, dispersed viewing pattern when looking at abstract pictures and a more local viewing pattern when looking at realistic pictures. Strikingly, artists generally looked in a more global way, fixating less on informative objects and more on areas holding no objects (e.g. water surface) compared with laypeople (Vogt & Magnussen, 2007).

This difference between more global and more local viewing patterns was also reported (although not quantified) by O'Neill and colleagues (2011), who investigated viewing patterns of subspecialists and trainees in ophthalmology. Inspection of gaze data showed that glaucoma subspecialists 'adapted' their viewing behavior more to the type of glaucomatous damage (focal loss or diffuse loss), whereas residents did not show adaptation: They showed more focal viewing behavior for both types of diseases. The 'adaptation' of viewing behavior to image type can be considered a sign of selective information processing, which is in line with Haider and Frensch's information reduction theory. Adaptation of viewing behavior to the type of image reflects selective processing: A distinction is made between relevant and irrelevant information and mainly relevant information is taken in. Selective processing lowers the amount of information that is to be processed in working memory, thereby lowering cognitive load.

We systematically explored the bottom-up effects of type of image on eye-movement patterns of radiologists, residents and students who diagnosed conventional chest radiographs (X-ray

images), on the basis of expertise theories of Kundel *et al.* (2007), Ericsson and Kintsch (1995) and Haider and Frensch (1999). Kundel's model predicts larger average saccade length for experts compared with students and residents. Ericsson and Kintsch's theory predicts shorter average fixation durations with increased expertise. Additionally, lower average fixation durations are expected on normal images in comparison with images showing a disease (Manning *et al.*, 2006). Furthermore, we suggest that Haider and Frensch's theory predicts that experts adapt their viewing behavior better to the type of diseases: We expect a low global/local ratio for focal diseases, a higher ratio for diffuse diseases and an even higher ratio for normal images. These differences are expected to be strongest for radiologists and weakest for students.

METHODS

Participants

Novices were 11 sixth year medical students (4 male, 7 female), mean age 25.2 years ($SD = 1.1$). Intermediates were 10 residents (4 male, 6 female), mean age 30.4 years ($SD = 3.5$). Experts were nine radiologists, (7 male, 2 female), mean age 44.7 years ($SD = 9.1$).

Residents reported working for 45.4 hours per week on average ($SD = 6.2$). Radiologists reported working for 48.4 hours per week on average ($SD = 15.5$). Residents were asked to indicate for how many months they had worked as a resident. The average was 28 months ($SD = 22.4$ months). For radiologists, the average length of their career after board licensing was 15.6 years ($SD = 8.2$ years), the minimum was 6 years and the maximum was 27 years. Students' experience

with thorax radiographs was estimated in hours. The median experience was 6 hours.

Design

The design was a 3×3 mixed factorial design, with expertise (student, resident and radiologist) as a between-subjects variable and type of image (focal, diffuse and normal) as a within-subjects variable. Dependent variables were percentage correct diagnoses, trial duration, average fixation duration, average saccadic amplitude and global/local ratio.

Apparatus and materials

The study was conducted using a remote high-speed eye-tracker with a 500 Hz sampling rate (Eyelink 1000). Participants' head movements were restricted only by using a forehead rest, which allowed participants to speak. Images were presented on a 19-inch LCD screen with a resolution of 1024×768 pixels. Participants' utterances were recorded using a digital recorder.

Materials were 24 conventional PA (Posterior-Anterior) chest radiographs (X-rays). Sixteen images showed a disease, eight of which were focal diseases and eight were diffuse diseases. Focal images contained one or more abnormalities at a specific location, while the rest of the lung was not affected. In diffuse images, all lobes of both lungs were affected. Normal images showed no abnormalities. All disease images and two of the normal images were chosen to be slightly challenging even for radiologists to ensure sufficient time for inspection of the images.

Procedure

Before the start of the experiment, participants were asked to sign an informed consent form and report on their experience in radiology based on a small questionnaire. Participants were assured that participation was anonymous, and that data were used only for the purpose of this research. It was explained that they were about to inspect chest radiographs, and it was stressed that they should act as they would do in normal practice. Eye dominance was assessed using the Miles test (Miles, 1930). Images were presented in two blocks of 12 images. Calibration of the eye-tracker is required to estimate the gaze direction of participants. A 9-point calibration procedure (Holmqvist et al., 2011) was conducted before each block of 12 images. The participants were instructed to carefully look at nine circles appearing one by one on the screen. After that, they proceeded to the actual task. They were asked to orally provide only the most likely diagnosis for all images. To continue to the next image, they had to hit a button. They were asked to work as quick and as accurate as possible. Before both blocks, one image with an obvious abnormality (large pneumothorax) was shown for practice, and participants received feedback before continuing to the first trial. Participants viewed each image in its entirety; zooming was not possible.

Analyses

The global/local ratio was computed by dividing the number of long saccades ($>1.6^\circ$ of visual angle) by the number of short saccades ($<1.6^\circ$ of visual angle) (Zangemeister et al.,

1995). Because this ratio was skewed, a lognormal transformation was performed. The skewness improved from 6.9 to 0.6. Multilevel analysis was conducted for all dependent variables, random intercepts were allowed on the participant level and on the item level. For all analyses, the model with random intercepts had a significantly better fit than models that did not include random intercepts. The correct answer for each image was decided by an expert radiologist, based on the given image as well as other information available to this radiologist (diagnosis, CT-scans, lateral chest radiographs and follow-up). Furthermore, it was decided which other valid conclusions might be drawn based on each image. All data were scored by two independent scorers. Answers for one of the focal images were excluded from analysis because no consensus on correct or incorrect answers could be reached. For all other images, 94.5% were assigned the same score by the two scorers. Other scores were discussed until consensus was reached.

RESULTS

Diagnostic ability

For proportion correct diagnoses, a model that included random intercepts on participant level and item level, fixed effects for image and expertise, and also an interaction between image and expertise (Model 2) had a significantly better fit with the data than a model that only included random intercepts for participant and item (Model 1), $\chi^2(8) = 50.117$, $p < .001$. For total trial duration, a model that included fixed effects for image and expertise level as well as random intercepts for participant and item (Model 2) gave a significantly better fit than a model that only included random intercepts for participant and trial (Model 1), $\chi^2(4) = 12.97$, $p = .011$. A model that also included a fixed interaction effect between image and expertise level (Model 3) did not lead to a better fit compared with Model 2, $\chi^2(4) = 1.97$, $p = .74$, so Model 2 was used.

The results of the multilevel analyses for percentage correct and total trial duration are presented in Table 1. The intercept of .397 refers to the average proportion correct for radiologists on the focal items. The score for diffuse images is on average .048 higher, a non-significant difference. Thus, focal and diffuse images seemed equally difficult to diagnose, and any differences in eye movements between those two image types are not the result of differences in difficulty. Unavoidably, normal images were easier than disease images, and the proportion correct for normal images was significantly higher ($b = .332$) than the proportion correct for focal diseases. Furthermore, trial duration for normal images was shorter ($b = -8206$ ms) than the average trial duration of focal images ($b = 29396$ ms), although this difference was only marginally significant. A significant difference between students and radiologists was found for the proportion correct and trial duration; students' proportion correct diagnosis was on average .306 lower, and their trial duration was on average 18418 ms longer. No significant differences between radiologists and residents were found for those two variables. A significant interaction was found, indicating that students scored relatively high on the normal items ($b = .173$).

Table 1. Results of multilevel analysis: Proportion correct and trial duration by type of image and expertise level

Parameter	Proportion correct				Trial duration (ms)			
	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Intercept	.397***	.100	32.9	3.97	29396***	5347	48.9	5.50
Image = normal	.332**	.132	31.9	2.51	-8206*	4156	23.5	1.98
Image = diffuse	.048	.140	31.9	0.34	177	4429	23.6	0.04
Expertise level = student	-.306***	.065	194.9	4.71	18418***	5992	29.9	3.07
Expertise level = resident	-.068	.066	194.9	1.03	9750	6126	29.9	1.59
Interaction: normal * student	.173**	.083	637.8	2.09				
Interaction: normal * resident	.073	.085	637.8	0.86				
Interaction: diffuse * student	-.061	.088	637.8	0.69				
Interaction: diffuse * resident	-.048	.090	637.8	0.53				

b, regression coefficient; *SE*, standard error of *b*; *t*, student *t* statistic of the test against *b*=0. All tests are conducted against the reference group: focal images, experts.

**p* < .1,

***p* < .05,

****p* < .005

Average fixation duration

For average fixation duration, a better fit was found for a model that included random intercepts for participants and trials as well as fixed effects for image and expertise and an interaction, compared with a model that only included random intercepts, $\chi^2(8)=31.34$, $p < .001$. Table 2 shows the results of this multilevel analysis. The intercept of 294.0 ms refers to the average fixation duration of an expert on focal items. The average fixation duration for normal items was on average 28.3 ms shorter, the average fixation duration for diffuse items was on average 15.4 ms shorter in comparison with focal items, both differences were significant. Expertise effects were not significant, students' average fixation duration was on average only 1.3 ms longer, and residents' average fixation duration was on average 17.4 ms longer. One interaction effect was found, indicating that students have relatively higher average fixation duration for normal images, ($b = 18.2$ ms).

Saccadic amplitude

Model 1 included random intercepts on participant level and on item level. Model 2 additionally included fixed effects for type of image and for expertise and an interaction between those two. Model 2 had a significantly better fit to the data,

$\chi^2(8)=46.52$, $p < .001$. Table 2 shows the results of this multilevel analysis. No significant effects of expertise were found, but significant effects of image were present. The intercept of 3.77 refers to the average saccadic amplitude for experts on focal images. Average fixation durations were on average 0.66° longer on normal images and 0.53° longer on diffuse diseases. Furthermore, a significant interaction was found indicating that students' saccadic amplitude was on average 0.44° shorter compared with radiologists on the normal images.

Global/local ratio

Model 1 included random intercepts on participant level and item level, Model 2 also included fixed effects of type of image and expertise and an interaction between type of image and expertise. Model 2 had a significantly better fit to the data compared with Model 1, $\chi^2(8)=27.27$, $p < .001$. Table 3 shows results of the multilevel analysis of the global/local ratio. The intercept of 1.14 refers to the average log-normalized global/local ratio of radiologists for focal images. A significant difference in the global–local ratio was found for normal compared with focal diseases ($b=0.45$), the global/local ratio was highest for normal images. The global/local ratio was marginally significantly higher for diffuse compared with focal

Table 2. Results of multilevel analysis: Average fixation duration and average saccadic amplitude by type of image and expertise level

Parameter	Average fixation duration (ms)				Average saccadic amplitude (degrees)			
	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i>	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Intercept	294.0***	11.2	38.7	26.3	3.77***	0.23	53.8	16.3
Image = normal	-28.3***	5.9	39.5	4.8	0.66***	0.20	30.3	3.3
Image = diffuse	-15.4**	6.3	39.7	2.5	0.53**	0.21	30.4	2.5
Expertise level = student	1.3	14.3	32.3	0.1	0.13	0.26	34.1	0.5
Expertise level = resident	17.4	14.7	32.3	1.2	0.20	0.26	34.2	0.8
Interaction: normal * student	18.2**	4.6	659.0	4.0	-0.44***	0.11	651.0	4.0
Interaction: normal * resident	6.6	4.7	659.1	1.4	0.14	0.11	651.0	1.3
Interaction: diffuse * student	6.5	4.9	659.1	1.3	-0.05	0.12	651.0	0.4
Interaction: diffuse * resident	5.6	5.0	659.1	1.1	0.11	0.12	651.0	0.9

b, regression coefficient; *SE*, standard error of *b*; *t*, student *t* statistic of the test against *b*=0. All tests are conducted against the reference group: Focal images, experts.

***p* < .05,

****p* < .005

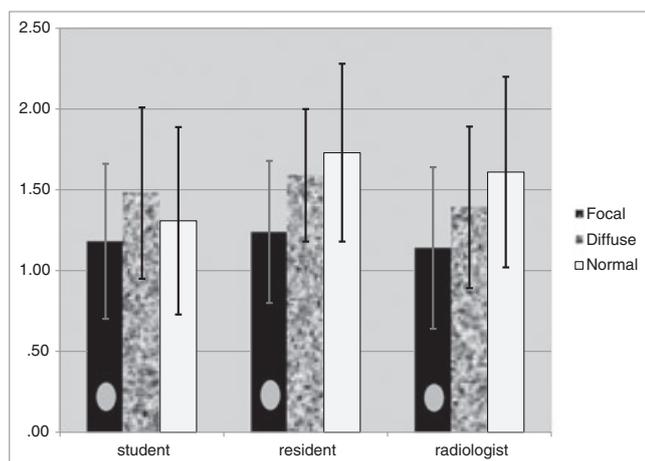


Figure 2. Normalized G/L ratio. Error bars reflect standard deviations

diseases ($b=0.25$). No significant effects of expertise were found. An interaction was found, which indicated that for normal images, the students had a significantly lower G/L ratio compared with radiologists ($b=-0.32$) see Figure 2.

An example of a typical focal pattern (low G/L ratio) and a typical diffuse pattern (high G/L ratio) can be found in Figure 3: The focal pattern means that in one location, many fixations are close to each other. In the typical diffuse pattern, there is not one location in which clustering of fixation takes place. The fixations are further apart. Figure 4 illustrates viewing patterns of students, residents and radiologists for the three types of images. It can be seen that for the focal image, one location (the location of the abnormality) specifically attracts attention. This is not the case for diffuse and normal images. Here, attention is more globally distributed over the image.

DISCUSSION

We investigated bottom-up effects of type of image on viewing patterns of students, residents and radiologists. Type of image played an important role in the way they looked at radiological

Table 3. Results of multilevel analysis: Log normalized global/local ratio

Parameter	Global/local ratio			
	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i>
Intercept	1.14***	0.18	50.3	6.4
Image = normal	0.45***	0.14	40.2	3.2
Image = diffuse	0.25*	0.15	40.3	1.7
Expertise level = student	0.03	0.22	36.1	0.2
Expertise level = resident	0.10	0.22	36.1	0.5
Interaction: normal * student	-0.32**	0.11	651.1	2.9
Interaction: normal * resident	0.01	0.11	651.1	0.1
Interaction: diffuse * student	0.09	0.12	651.1	0.7
Interaction: diffuse * resident	0.10	0.12	651.1	0.9

b, regression coefficient, *SE*, standard error of *b*, *t*, student *t* statistic of the test against $b=0$. All tests are conducted against the reference group: focal images, experts.

* $p < .1$,

** $p < .05$,

*** $p < .005$

images. Regardless of expertise, participants looked relatively long at specific locations in focal images, whereas in diffuse images, a more dispersed pattern was found, with lower average fixation durations. Saccadic amplitudes were highest for normal images, slightly lower for diffuse images and lowest for focal images. Students showed different viewing behavior for normal images: Their average fixation durations were higher than those of residents and radiologists, whereas their saccadic amplitudes and global/local ratio were lower. In conclusion, differences in viewing patterns between experts and novices are relatively small on disease images but larger on normal images. In contrast, students' diagnostic accuracy on normal images is relatively high, whereas for disease images their accuracy is very low.

Predictions about expertise differences on the different measures were made based on the holistic model of image perception (Kundel et al., 2007), the theory of long-term working memory (Ericsson & Kintsch, 1995) and information reduction theory (Haider & Frensch, 1999). The holistic model of image perception predicts that with higher expertise, longer average saccade lengths would be found. The theory of long-term working memory predicts shorter average fixation durations with increased expertise. Both expertise effects were not found, neither for the diffuse nor for the focal diseases. Rather, strong effects of type of image were found, comparable with the ones we expected for the global/local ratio. Not only did radiologists adapt their viewing patterns but also residents and even students changed their viewing patterns according to the different types of diseases.

Haider and Frensch's information reduction theory predicts more selective information processing with increased expertise: Experts have a better ability to distinguish relevant from irrelevant information, and they selectively take in mainly relevant information. This leads to lower cognitive load for experts compared with novices. The effects of the type of image on the average fixation duration and saccade length as well as the global/local ratio seem to indicate that selective processing has taken place. In a focal disease, most information can be gathered when one specific location is inspected in depth (i.e. a low global/local ratio and longer average fixation duration). For diffuse diseases, more information is gained from examining together the elements that make up the pattern. This requires a higher global/local ratio and shorter average fixation durations. For normal images, even shorter average fixation durations are required. It was shown before in research with chest radiographs that average fixation duration on true positives (correctly diagnosing a lung nodule) is longer than average fixation duration on true negatives (correctly diagnosing normality) (Manning et al., 2006).

In contrast to the predictions of Haider and Frensch's theory, we found that the viewing behavior of students was quite similar to that of residents and radiologists. Students adapted their viewing pattern to the type of image for focal and diffuse diseases. But although students' viewing behavior was similar to that of residents and radiologists, they did not perform nearly as well as the two other groups, consistent with the results of Crowley, Naus, Stewart, & Friedman (2003) and Mello-Thoms et al. (2012). Both report that perceptual aspects of the task (i.e. *detecting* abnormalities) were developed before participants were able to correctly *interpret* the abnormalities

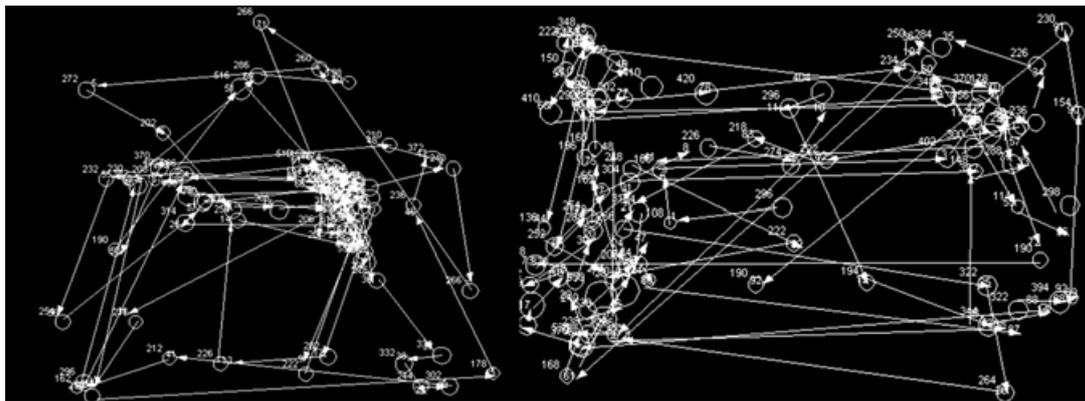


Figure 3. Eye tracking data of one participant showing a typical focal pattern (left) and typical diffuse pattern (right)

	Students (n=11)	Residents (n=10)	Radiologists (n=9)
Focal			
Diffuse			
Normal			

Figure 4. Duration-based heat maps for a focal disease, a diffuse disease and normal image. Data are aggregated per group

and to integrate these in a correct diagnosis. Lesgold et al. (1988) posed a similar developmental trajectory, in which perceptual processing develops before cognitive processing. The students in our sample showed differentiation in fixation duration, saccadic amplitude and global/local ratio between the two types of disease, indicating perceptual development. However, they were not yet able to interpret their findings into a diagnosis. Furthermore, students were able to decide for most of the normal images that no abnormality was present. This suggests that students' selective processing is largely driven by bottom-up, stimulus-level differences between the three types of

images rather than being based on knowledge. Selecting mainly the relevant information did not yet allow students to correctly interpret this information into a correct diagnosis, as they did not yet possess the necessary knowledge for that.

In contrast to their relatively high performance on normal images, expected eye movement differences between experts and novices were found for the normal images: Students had significantly higher average fixation duration and significantly lower average saccadic amplitude compared with radiologists. Their global/local ratio was also significantly lower than the ratio of experts.

Some limitations are worth noting. The number of participants in this study was quite small, so that a replication with a larger sample is desirable. Furthermore, the normal diagnostic process slightly differs from our experimental set-up. Specifically, radiologists usually view a radiological image in combination with a patient history including patient complaints and specific questions, while in this study, we required participants to give a diagnosis based on a chest radiograph only. It is commonly found (Berbaum, Franken, Caldwell, & Scharzt, 2010) that patient history and clinical information influence not only performance but also viewing behavior. In this study, we were mainly interested in visual processes. As these processes could be influenced by patient history in a manner that is hard to control, we decided not to provide participants with any information additional to the radiographs, and to analyze how they diagnosed based on these radiographs only. Finally, the current study did not test a specific educational intervention, but compared characteristics of 6th year students, residents and radiologists. Thus, conclusions regarding the development of learning materials or course design cannot be made based on the results. However, it seems that further research on educational interventions could focus on teaching students not only interpretation skills but also detection based on elaborate knowledge.

In this study, we investigated viewing patterns of different expertise groups in terms of fixation durations and distribution of fixations. Our approach of using a global/local ratio is a novel way of investigating top-down expertise differences in visual diagnostic reasoning, when analyzing areas of interest (Holmqvist et al., 2011) is not possible. Its advantage is that it also considers bottom-up effects of type of image. This novel approach for investigating visual expertise could also be relevant for other domains of visual expertise, such as ophthalmology (O'Neill et al., 2011), art (see Vogt & Magnussen, 2007; Zangemeister et al., 1995) or meteorology (Hegarty, Canham, & Fabrikant, 2010).

Although for disease images, the viewing patterns of students are reasonably similar to those of radiologists, their diagnostic ability is much lower. Radiologists possess a large body of knowledge that helps them correctly make diagnostic decisions based on the information that they perceive (Wood, 1999). Although students might look in ways comparable with those of radiologists and residents, they still lack this body of knowledge that helps them to interpret the visual information and give the correct diagnosis. Claudia Mello-Thoms et al. (2008) stated that radiology-experts have 'different search patterns and similar decision outcomes' (p. 212.). The opposite seems true when experts and novices are compared: Similar search patterns lead to different decision outcomes.

ACKNOWLEDGEMENTS

The authors would like to thank Halszka Jarodzka for her advice regarding the eye tracking methodology, Jessica Werthmann for enabling the use of the eye tracking equipment and Jimmie Leppink for advice regarding the multilevel analyses. Furthermore, the authors would like to thank two anonymous reviewers for their very helpful feedback.

REFERENCES

- Berbaum, K. S., Franken, E., Caldwell, R. T., & Scharzt, K. M. (2010). Satisfaction of search in traditional radiographic imaging. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 107–138). New York: Cambridge University Press.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*, 20, 155–166.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55–81.
- Crowley, R. S., Naus, G. J., Stewart, J., & Friedman, C. P. (2003). Development of visual diagnostic expertise in pathology: An information-processing study. *Journal of the American Medical Informatics Association*, 10, 39–51.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Ericsson, K. A., Charness, N., Feltovich, P., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23, 523–552.
- Groen, G. J., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 287–309). Hillsdale, NJ: Erlbaum.
- de Groot, A. D. (1946). Het denken van den schaker: Een experimenteel-psychologische studie. Noord-Hollandsche Uitgevers Maatschappij.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 172–190.
- Hansell, D. M., Bankier, A. A., MacMahon, H., McLoud, T. C., Mueller, N. L., & Remy, J. (2008). Fleischner society: Glossary of terms for thoracic imaging. *Radiology*, 246, 697–722.
- Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 37–53.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. New York: Oxford University Press.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203.
- Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20, 146–154.
- Kok, E. M., De Bruin, A. B. H., Robben, S. G. F., & Van Merriënboer, J. J. G. (in press). Learning radiological appearances of diseases, does comparison help? *Learning and Instruction*. Advance online publication. doi: 10.1016/j.learninstruc.2012.07.004.
- Krupinski, E. A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic Radiology*, 12, 965–969.
- Krupinski, E. A. (2010). Perceptual factors in reading medical images. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 81–90). New York: Cambridge University Press.
- Krupinski, E. A., & Roehrig, H. (2010). Optimization of display systems. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 395–405). New York: Cambridge University Press.
- Krupinski, E. A., Berger, W. G., Dallas, W. J., & Roehrig, H. (2003). Searching for nodules: What features attract attention and influence detection? *Academic Radiology*, 10, 861–868.
- Kundel, H. L., & Nodine, C. F. (1983). A visual concept shapes image perception. *Radiology*, 146, 363–368.
- Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, 242, 396–402.
- Leong, J. J. H., Nicolaou, M., Emery, R. J., Darzi, A. W., & Yang, G. Z. (2007). Visual search behaviour in skeletal radiographs: A cross-specialty study. *Clinical Radiology*, 62, 1069–1077.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

- Manning, D. J., Ethell, S. C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *British Journal of Radiology*, *77*, 231–235.
- Manning, D. J., Ethell, S. C., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography*, *12*, 134–142.
- Mello-Thoms, C., Ganott, M., Sumkin, J., Hakim, C., Britton, C., Wallace, L., & Hardesty, L. (2008). Different search patterns and similar decision outcomes: How can experts agree in the decisions they make when reading digital mammograms? In E. Krupinski (Ed.), *Digital mammography* (Vol. 5116, pp. 212–219). Berlin/Heidelberg: Springer.
- Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., Stalder, J., & Maitz, G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology*, *12*, 830–840.
- Mello-Thoms, C., Mello, C. A. B., Medvedeva, O., Castine, M., Legowski, E., Gardner, G., Tseytlin, E., & Crowley, R. S. (2012). Perceptual analysis of the reading of dermatopathology virtual slides by pathology residents. *Archives of Pathology & Laboratory Medicine*, *136*, 551–562.
- Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology*, *3*, 412–430.
- Nodine, C., & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 139–156). New York: Cambridge University Press.
- Norman, G. R., Coblenz, C. L., Brooks, L. R., & Babcock, C. J. (1992). Expertise in visual diagnosis – a review of the literature. *Academic Medicine*, *67*, S78–S83.
- O'Neill, E. C., Kong, Y. X. G., Connell, P. P., Ong, D. N., Haymes, S. A., Coote, M. A., & Crowston, J. G. (2011). Gaze behavior among experts and trainees during optic disc examination: Does how we look affect what we see? *Investigative Ophthalmology & Visual Science*, *52*, 3976–3983.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.
- Remington, R. W., Johnston, J. C., Ruthruff, E., Gold, M., & Romera, M. (2000). Visual search in complex displays: Factors affecting conflict detection by air traffic controllers. *Human Factors*, *42*(3), 349–366.
- Samei, E., Flynn, M. J., Peterson, E., & Eyler, W. R. (2003). Subtle lung nodules: Influence of local anatomic variations on detection. *Radiology*, *228*, 76–84.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review*, *5*, 205–221.
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise – theory and implications. *Academic Medicine*, *65*, 611–621.
- Van Gog, T., Ericsson, K. A., Rikers, R. M. P. J., & Paas, F. (2005). Instructional design for advanced learners: Establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educational Technology Research and Development*, *53*, 73–81.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*, 147–177.
- Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*, 91–100.
- Wood, B. P. (1999). Visual expertise. *Radiology*, *211*, 1–3.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, *33*, 1009–1025.