

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Chen, W;HolcDorf, D;McCusker, MW;Gaillard, F;Howe, PDL

Title:

Perceptual training to improve hip fracture identification in conventional radiographs

Date:

2017-12-01

Citation:

Chen, W., HolcDorf, D., McCusker, M. W., Gaillard, F. & Howe, P. D. L. (2017). Perceptual training to improve hip fracture identification in conventional radiographs. Plos One, 12 (12), <https://doi.org/10.1371/journal.pone.0189192>.

Persistent Link:

<https://hdl.handle.net/11343/255537>

License:

[CC BY](#)

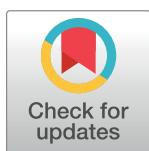
RESEARCH ARTICLE

Perceptual training to improve hip fracture identification in conventional radiographs

Weijia Chen^{1*}, David HolcDorf², Mark W. McCusker², Frank Gaillard^{2,3}, Piers D. L. Howe¹

1 School of Psychological Sciences, University of Melbourne, Parkville, Australia, **2** Radiology Department, Royal Melbourne Hospital, Parkville, Australia, **3** Radiology Department, University of Melbourne, Parkville, Australia

* weijiacelinachen@gmail.com



Abstract

Diagnosing certain fractures in conventional radiographs can be a difficult task, usually taking years to master. Typically, students are trained ad-hoc, in a primarily-rule based fashion. Our study investigated whether students can more rapidly learn to diagnose proximal neck of femur fractures via perceptual training, without having to learn an explicit set of rules. One hundred and thirty-nine students with no prior medical or radiology training were shown a sequence of plain film X-ray images of the right hip and for each image were asked to indicate whether a fracture was present. Students were told if they were correct and the location of any fracture, if present. No other feedback was given. The more able students achieved the same level of accuracy as board certified radiologists at identifying hip fractures in less than an hour of training. Surprisingly, perceptual learning was reduced when the training set was constructed to over-represent the types of images participants found more difficult to categorise. Conversely, repeating training images did not reduce post-training performance relative to showing an equivalent number of unique images. Perceptual training is an effective way of helping novices learn to identify hip fractures in X-ray images and should supplement the current education programme for students.

OPEN ACCESS

Citation: Chen W, HolcDorf D, McCusker MW, Gaillard F, Howe PDL (2017) Perceptual training to improve hip fracture identification in conventional radiographs. PLoS ONE 12(12): e0189192. <https://doi.org/10.1371/journal.pone.0189192>

Editor: Etsuro Ito, Waseda University, JAPAN

Received: August 10, 2017

Accepted: November 21, 2017

Published: December 21, 2017

Copyright: © 2017 Chen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by a Royal Australian and New Zealand College of Radiology (RANZCR) research grant in 2015 to FG and PDLH (no grant number available, URL <https://www.ranzcr.com/college/awards-and-prizes/research-awards-and-grants>), and the University of Melbourne Engagement Grants Scheme 2015 to FG (no available grant number or URL). The funders had no role in study design, data collection

Introduction

Diagnosing medical imagery is a complex task. To identify abnormalities and reach a correct diagnosis, radiologists must correctly interpret a large number of visual cues. These cues can be masked by the generally noisy background in an X-ray image, variation between patients, slight differences in positioning, or even be mistaken for normal anatomical structures [1]. Residents are typically trained to diagnose conventional radiography in a rule-based manner, usually with no effort made to cluster numerous examples of the same study or provide immediate gold-standard feedback. Given the visual complexity of X-ray images, learning to diagnose radiography in this manner can be difficult. In this study, we investigated how perceptual learning may apply to radiology training and whether students could more rapidly learn to diagnose proximal neck of femur fractures in conventional frontal radiographs of the hip.

Humans can easily make perceptual distinctions that are hard to verbalise. A classic example is sex identification. People usually have no trouble correctly recognising a face as male or

and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Dr. Gaillard is the founder/CEO of Radiopaedia.org, outside the submitted work. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

female and rely on a large number of visual cues to do so. These cues include distances between key points on the face [2], the luminance contrast between the eyes, lips, and the surrounding skin [3], the shape of facial features [4], and hair cues [5]. Although each of these cues is individually unreliable, people are able to combine them to make rapid and robust decisions [2, 3, 6]. Traditionally, advanced visual recognition skills were assumed to develop only from years of practice because they are difficult to explicitly teach. Recent studies, however, suggest that such skills can in fact be acquired rapidly within days, or even hours, via perceptual training [7].

The term *perceptual training* refers to any training whose goal is to improve perceptual skills, such as the ability to recognise or categorise images. The learning that results from such training is known as *perceptual learning* [1, 8]. Instead of learning to interpret a medical image by following a set of explicit rules, students can instead be trained to visually recognise abnormalities. While there have been a large number of perceptual training studies (for a review see [9]), to the best of our knowledge, there have been only three studies that investigated perceptual training in the context of diagnosing medical images, and only one in radiology. Sowden, Davies and Roling [1] trained novices to identify microcalcification clusters in mammograms by presenting them with 60 images, three times a day. Each image contained a microcalcification cluster and the task was to localise it. Immediate feedback was given. An improvement of 17% in localisation accuracy was observed after three days of training. Similarly, Krasne, Hillman, Kellman and Drake [10] trained first and second year medical students in skin histopathology using 261 unique images. The students were required to classify each image as containing pathological or normal histology. The students were immediately informed if they were correct or not. Despite a median training time of only 15 minutes, mean categorisation accuracy increased by approximately 13%. Finally, in a recent study by Xu, Rourke, Robinson and Tanaka [11], naïve undergraduate students were shown 100 images, one at a time, and asked to indicate whether or not the image contained a melanoma. As before, the students were immediately informed if they were correct or not. When subsequently tested on a new set of images, their false alarm rate had decreased by 19% compared to their pre-training performance.

Much of the previous research into perceptual learning has assumed that it is a low-level phenomenon that is achieved by the modification of receptive fields of neurons in early sensory cortical areas that are responsible for the initial encoding of the stimulus [1, 9]. However, accumulating evidence has argued against this viewpoint. In particular, it has been found that perceptual learning will sometimes transfer between retinal locations [7, 12]. This finding is not compatible with the assumption that perceptual learning is achieved by neurons in early sensory cortical areas as these neurons have highly localised receptive fields. It has also been reported that the changes in receptive field structure in early cortical areas are often too small to account for the behavioural changes induced by perceptual learning [13]. More robust receptive field changes seem to occur at later stages of visual processing, for example cortical area V4 [13].

For the purposes of this study, we will assume that perceptual learning is not achieved solely by changing the receptive fields in early cortical areas. Instead, following the lead of Eleanor Gibson [8], we will conceptualise it terms of later cortical areas learning to select the outputs of earlier cortical areas that are most relevant for making the required perceptual distinction. Following Petrov et al., we assume that the selection procedure is not all or nothing but instead occurs by later cortical areas weighting the outputs of early cortical areas when making decisions [14]. For conceptual simplicity, we think of this process as being approximately by a multilayer perceptron [15]. A multilayer perceptron is an abstract representation of the neural system. It is an artificial neural network comprising multiple layers of nodes, each layer

typically fully connected to the next one as a directed graph. Each node performs a weighted sum of its inputs and passes this value through a nonlinear activation function to determine its output. The weights therefore determine the degree to which nodes in one layer can affect the activities of nodes in a subsequent layer. Perceptual learning can be achieved by altering these weights so as to select the outputs of the previous layer that are most relevant for making the required perceptual distinction. As such, multilayer perceptrons represent a useful and practical model of perceptual learning [15] and can be used to make novel predictions.

In our study, we investigated perceptual training in an area in which it had not previously been applied, diagnosing radiographs. From the previous literature, it is clear that perceptual learning can improve a participant's ability to diagnose medical images. However, it is less clear to what extent this improvement is practically useful. For example, could perceptual training be used to train a naïve participant up to the level of a board certified radiologist or would the required amount of training be prohibitive? Our first aim was therefore to quantify the amount of perceptual training required to do achieve this level of performance.

The second aim of our study was to determine the best way to optimise our perceptual training. Multilayer perceptrons generally learn quicker when trained only on the more difficult training images [16]. To the extent that multilayer perceptrons model perceptual learning, it follows that perceptual learning should also occur quicker when humans are exposed only to the more difficult training images. Our second aim was to test this prediction.

Our final aim was to investigate to what extent images can be reused during perceptual learning. It has been shown the repeatedly showing the same training images is an effective way to train multilayer perceptrons [15]. Our third aim was therefore to investigate whether repeating image presentations resulted in further perceptual learning (as opposed to showing the images just once) and, if so, to compare the degree of perceptual learning achieved by repeating images to the degree of perceptual learning achieved by showing an equivalent number of novel images. It is often expensive and time-consuming to obtain medical images. Being able to reuse medical images would therefore make perceptual training more feasible.

Method

Participants

The previous study that is most similar to ours suggests that significant effects can be observed with just 10 observers [11]. Because we are in a different area (radiology as opposed to dermatology), as a precaution we used more participants. For the six experiments listed below, in total, 142 novices with no prior knowledge in X-ray film interpretation participated. The number of novices in each experiment is shown in Table 1. They were mostly undergraduate students from The University of Melbourne, aged between 17 and 50 years (M age = 23.5 years, SD = 5.5 years; 49 men). Additionally, we also evaluated the performance of board certified radiologists and first and second year radiology residents. Unlike the novices, these

Table 1. The number of participants in Experiments 1–6.

Experiment	No. of Participants
1	23
2	25
3	22
4	23
5	24
6	25

<https://doi.org/10.1371/journal.pone.0189192.t001>

participants received no training and did not participate in the experiments listed below. The five board certified radiologists were between the ages of 30 and 41 years (M age = 35.8 years, SD = 5.4 years; 4 men) and had an average experience of 8.5 years. The five 1st and 2nd year residents were between the ages of 25 and 32 years (M age = 28.0 years, SD = 3.2 years; 3 men) and had an average experience of 1.2 years. All participants had normal or corrected-to-normal visual acuity (at least 20/25; near-field Snellen eye chart) and normal colour vision (Ishihara plates). This study was approved by The University of Melbourne Human Research Ethics Committee (Ethics ID 1545113) and all experiments were performed in accordance with the ethical standards laid down in the amended Declaration of Helsinki [17]. The research ethics committee approved the lack of parent or guardian consent for participants under the age of 18. All participants gave written informed consent. Six experiments were conducted in total, from 19th January to 16th November, 2016. Those in the novice group were paid AU\$15/hour for their participation, and in Experiments 2–6 they could additionally earn a bonus of AU\$5 if they scored above 85% in the post-training test. Results were analysed using IBM SPSS Statistics 22.

Apparatus and stimuli

The stimuli were cropped hip region images from AP pelvis X-rays selected from the film archive of the Royal Melbourne Hospital of patients who presented to the emergency department with a fracture of head or neck of femur, in all cases confirmed surgically. The images were anonymized and digitised using the FujiFilm Synapse PACS v4.5 software and presented on a black background. Annotations were made using Skitch by Evernote. From each pelvis X-ray each hip region was independently cropped generating two non-overlapping images, one of the left hip and the other of the right hip, one of which would contain a fracture, while the other would not. For the sake of consistency, these images were always presented as the right hip (i.e., any left hip images were flipped so that they looked like a right hip). X-rays were excluded if the fracture occurred in a region with abnormal underlying bone (i.e. pathological fracture due to bone tumour/abnormal underlying bones abnormal, such need to include this—I' any metalware (i.e. total hip replacement, dynamic hip screws). X-rays were also excluded if the non-fractured hip exhibited grossly abnormal pathology, such as tumours or Paget disease. Osteoporosis and osteoarthritis in either hip were not considered to be exclusion criteria. Images were also excluded if there was any metalware (i.e. total hip replacement, dynamic hip screws) in either hip.

Procedure

The stimuli were presented on a personal computer using MATLAB[®] and the Psychophysics Toolbox [18, 19] subtending an area of $25.8^\circ \times 16.2^\circ$ at the 60 cm viewing distance. They were presented in darkened room similar to the conditions under which radiologists report. A single image was presented on each trial and participants were instructed to indicate whether the X-ray image contained a fracture or not by pressing one of two keys on the keyboard. The image remained on screen until a response was made.

Pre-training test. In Experiments 1 and 2, each participant in the novice group was first tested on 20 images to establish their pre-training accuracy. No feedback was given in this stage. There was no pre-training test for Experiments 3–6.

Training phase. The training phase was conducted in a single session in Experiments 1–5. In Experiment 1, participants were shown 200 unique images, one at a time, and asked to indicate whether or not the image contained a fracture. Participants received feedback on every trial. Specifically, if they made a mistake, an error message appeared and the same image

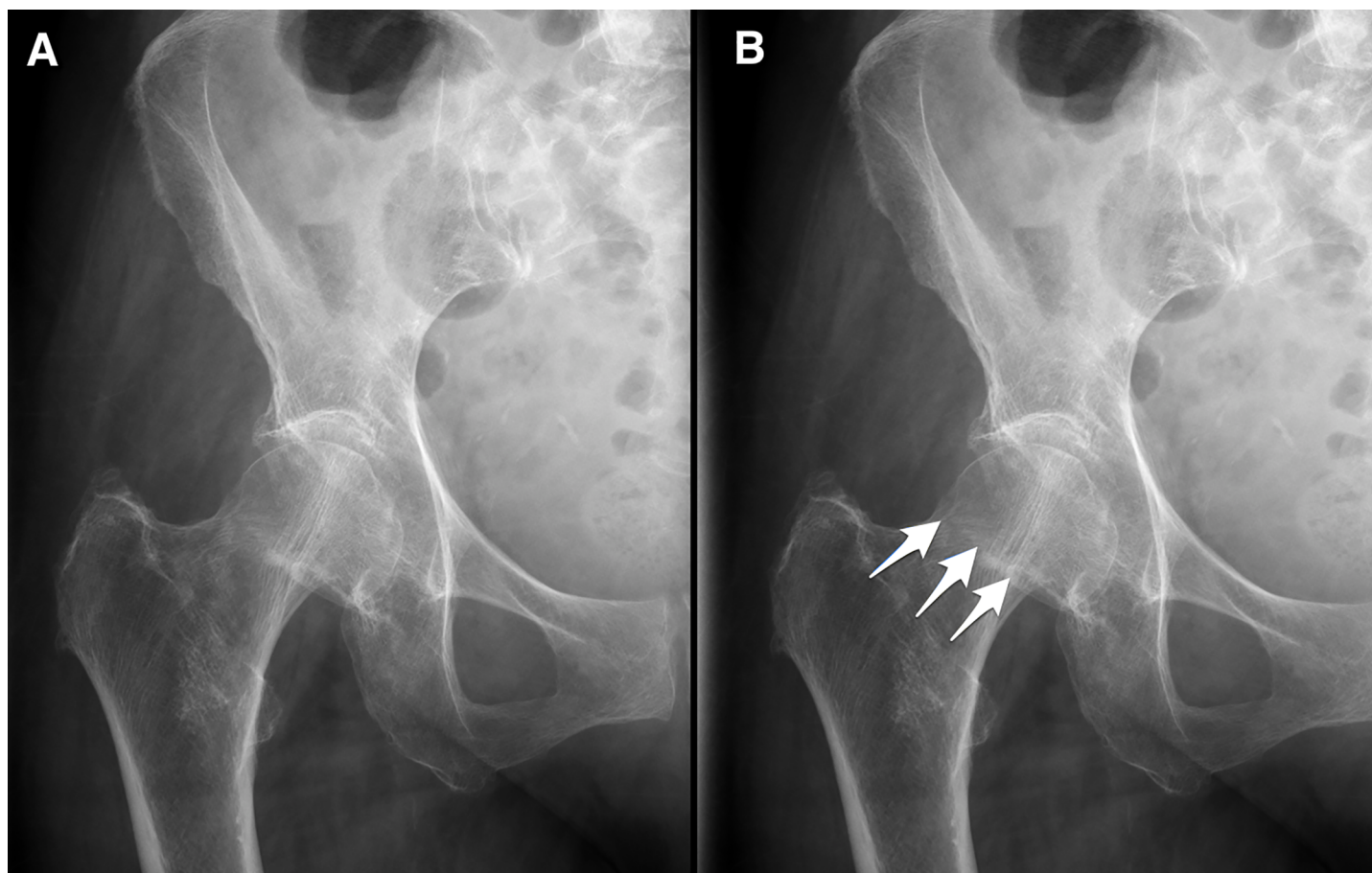


Fig 1. Examples of images shown to participants. (a) An example of the image initially shown to a participant. The image is of the right hip and pelvic region of a 76-year-old female. A fracture is visible. (b) An example of the feedback the participant would receive if the participant were, in this case, to incorrectly state that there was no fracture.

<https://doi.org/10.1371/journal.pone.0189192.g001>

was shown again, with arrows indicating the location of the fracture, if there was one (Fig 1). Participants earned one point for each correct response and the total score was shown on screen to motivate performance. In Experiment 2, the training procedure was the same but the novices were only trained on the relatively more difficult images based on the performance of the novice group in Experiment 1. This was to test the hypothesis that training solely on the more difficult images would lead to greater perceptual learning. Sixty-two images that had previously been classified correctly 5–75% of the time were selected as training images. We chose a wide accuracy range so that it was difficult but still possible to classify those images correctly. We purposely chose a large range of accuracies to minimise the possibility that we were only training observers to categorise images of a particular difficulty. This training block was presented to participants four times for a total of 248 trials. In Experiment 3, novices viewed a single block of 320 unique images. These images were selected randomly without regard to difficulty level. In Experiment 4, each participant was presented with the same block of 320 images twice; and in Experiment 5 participants viewed a single block of 640 unique images. This was to investigate the effect of repeating the training image set on post-training accuracy. Experiment 6 was conducted over two days. On the first day of training, each participant was presented with 640 randomly chosen X-ray images. They then viewed the same block of training images on the following day. In all six experiments, the target prevalence rate was 50%.

Post-training test. After the training session, each participant was tested on a single block of new images to assess their post-training accuracy. No feedback was provided in this test. The post-training test consisted of 180 images in Experiments 1 and 2. This was reduced to 80 images in Experiments 3–6, because novices continued to learn in the post-training test even in the absence of feedback. The training and test images were randomly selected from an image pool without replacement. The board certified radiologists and 1st and 2nd year residents received no training and were tested on 180 images randomly selected from the same image pool as was used for the novices.

Results

Accuracy was calculated as the sum of hits and correct rejections divided by the total number of trials. The data of nine participants from Experiments 1–5 were excluded from the analysis because their accuracy was at or below the chance level of 50% throughout the experiment, indicating a lack of engagement with the task. Two participants from Experiment 6 were excluded because they did not return on the second day of training.

In addition to the mean post-training accuracies for all novices, we also conducted analysis on the top five novices to explore how training affects the top quartile of our trainees. Because radiologists are selected on merit, we analysed the results of the top five performing novice participants in an attempt to control for ability variations, so as to give a more realistic estimate of the training effects. Participants in the novice group were ranked based on their accuracy in the first half of the post-training test. For each experiment, the top five performing participants were selected based on their responses to the first half the post-training test. We then calculated the mean accuracy of these participants based on their responses to the second half of the post-training test.

The average pre-training accuracy for the entire novice group obtained from Experiments 1 and 2 was 55.9%. This was significantly higher than the guess rate of 50%, $t(47) = 2.31$, $p = .03$. However, all the novices who participated in this study confirmed before commencing the experiment that they had no prior knowledge of X-ray film reading. On average, it took participants 2.77 seconds to read each image in the training phase. The mean reading time for the experts in the test phase was 4.62 seconds per image.

The mean accuracies for the entire novice group in each experiment are presented in Fig 2. For the sake of comparison, the mean accuracy of the residents is shown by the dashed red

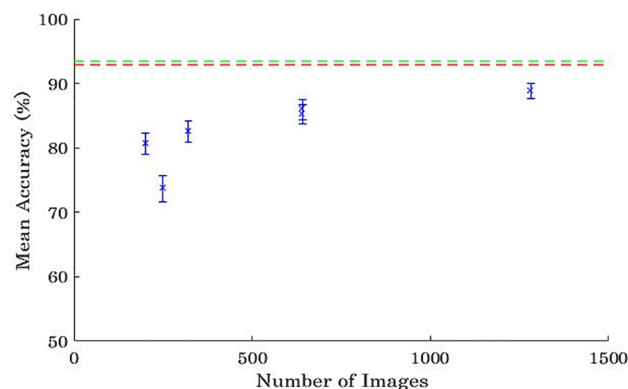


Fig 2. Mean accuracies of the board certified radiologists (dashed green line), radiology residents (dashed red line) and the average of all the undergraduates (blue data points) in Experiments 1–6 plotted against the number of training images viewed in each experiment. Error bars represent standard error of the mean.

<https://doi.org/10.1371/journal.pone.0189192.g002>

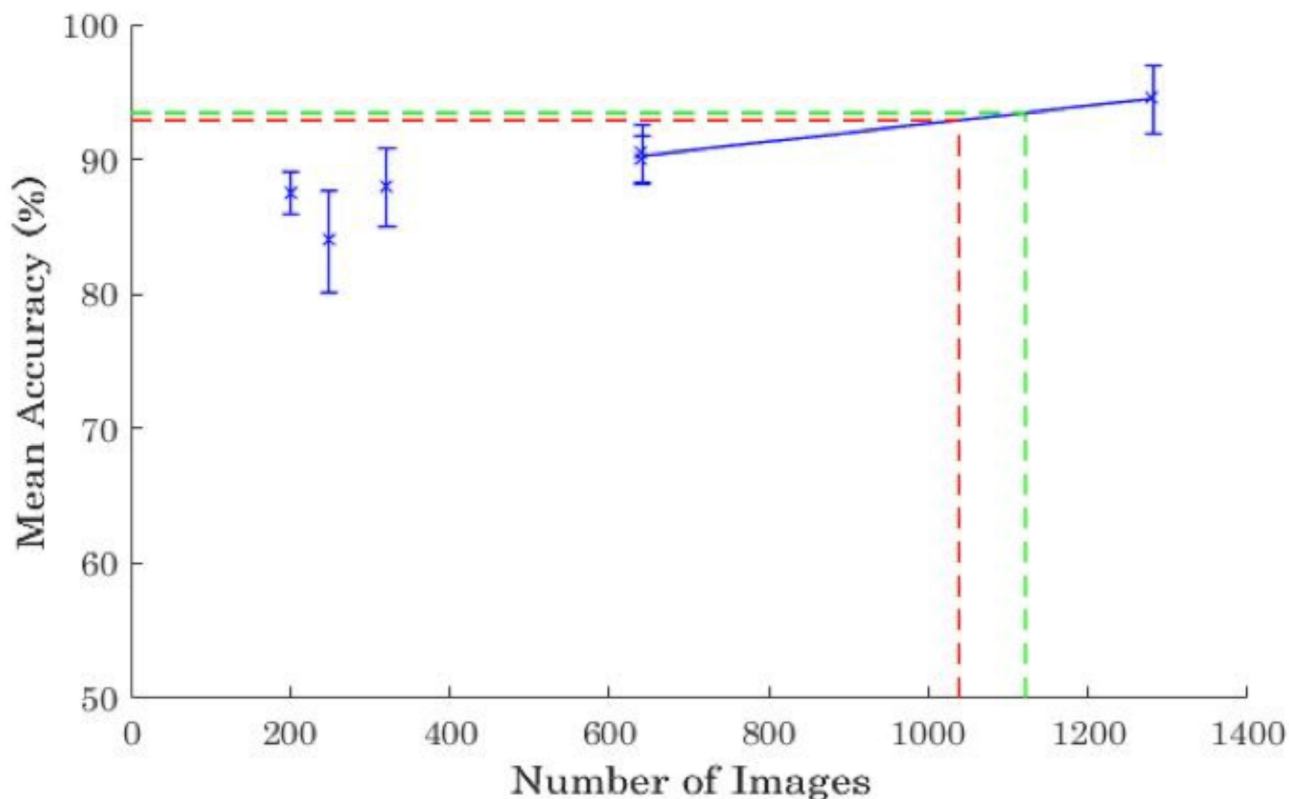


Fig 3. Mean accuracies of the board certified radiologists (dashed green line), radiology residents (dashed red line) and the top five undergraduates (blue data points) in Experiments 1–6 plotted against the number of training images viewed in each experiment. Error bars represent standard error of the mean. The data from Experiments 4–6 was interpolated to predict the number of training images required to bring the accuracies of the top five undergraduates up to the same level as the radiology residents and board certified radiologists.

<https://doi.org/10.1371/journal.pone.0189192.g003>

lines and the mean accuracy of the board certified radiologists are shown by the dashed green lines. Fig 3 shows the equivalent data for the top five participants for all the experiments. Interpolation was used to determine how many training images are needed to bring the level of the top five novices up to the level of the residents and board certified radiologists. In general, mean post-training accuracy increased as the number of training images increased, with the exception of Experiment 2. *t* tests revealed that the average post-training accuracy for all novices in Experiment 2 was lower than that in Experiment 1, $t(39) = 2.73$, $p = .01$, Cohen's $d = .85$, despite Experiment 2 utilising more training images than Experiment 1. This shows that training the participants on just the more difficult training images did not improve performance, contrary to what we predicted. Conversely, there was no significant difference in post-training accuracy in Experiments 4 and 5 regardless of whether the results are averaged over all novices ($t(43) = 0.32$, $p = .75$, Cohen's $d = .004$), or just over the top five novices ($t(8) = 0.18$, $p = .86$, Cohen's $d = .02$). This indicates that showing each image twice did not reduce the effectiveness of the training. As predicted, reusing images is an effective instructional strategy.

As shown by Fig 2, when accuracy is averaged over all novices, the mean accuracy of the novices did not exceed that of the residents or the radiologists, even when trained on 1280 images. While it is possible that further training might have improved performance further, it is unclear whether with sufficiently training could ever be brought up to the standard of residents. The performance of residents and radiologists may in part reflect the considerable amount of rule-based conceptual training they receive, in addition to perceptual training.

Conversely, Fig 3 shows that the mean accuracy of the top five novices exceeded that of the experts. While the underlying learning function for novices is likely to be a power function [20], it can be approximated by a linear function over a sufficiently narrow range. We therefore performed a linear interpolation of the data in Experiments 4–6 and predicted that training novices on 1,037 images could bring the mean accuracy of the top five novices to the same level of residents. To train on this number of images would take approximately 48 minutes. To increase the accuracy of top five novices to the same level of proficiency as board certified radiologists would require 1,121 images and approximately 52 minutes of training.

For conceptual ease, the above analysis was presented in terms of percentage correct. We obtained essentially the same results when we repeated this analysis using a bootstrap comparison [21] of d' (S2 File). The advantage of d' is that it is a measure of sensitivity that is independent of any response bias the observers may have [22]. In this case, discounting the response bias did not materially affect our results.

Discussion

The primary aim of this study was to determine how much perceptual training is required to bring naïve observers to the same level of proficiency as board certified radiologists in identifying hip fractures on frontal hip radiographs. It also explored ways to optimise the training procedure, and set a foundation upon which further research can build. For the average undergraduate in this study, training did not improve their accuracy to the same level as the experts. Moreover, the average reading time in our novice group was faster than the average time spent on each image by the experts. This pattern of response fits a classic speed-accuracy trade-off explanation for their performance. Biggs et al. [23] conducted a difficult conjunction search task with professional airport baggage screeners and non-professionals. They found that non-professional searchers terminated searches faster than professionals and that response time was the primary predictor of search accuracy for them. But the more experienced professionals responded slower and had more consistent RTs on every trial, in turn, RT consistency best predicted their performance. On the other hand, our results from the top five novices suggest that under one hour of training with approximately 1,100 images is sufficient to train the top performing undergraduates to the level of accuracy of board certified radiologists. This shows that perceptual training is highly effective and argues for its potential inclusion in the standard radiology curriculum, or at least for modification of existing ad-hoc training to emphasise clustering of study type and need for immediate feedback.

Comparing the findings of Experiments 1 and 2 reveals that training participants solely on the more difficult images can be detrimental to their performance compared to training them on a random sample of images. This was the opposite of what we predicted and may be due to the training sample in Experiment 2 (i.e. a sample containing only the more difficult images) being non-representative of the types of images radiologists generally see. The fact that no difference in post-training accuracy was observed between Experiments 4 and 5 demonstrate that repeating a training image does not necessarily reduce training effectiveness. This was expected and is in broad accordance with the multilayer perceptron literature where training typically involves repeatedly showing the multilayer perceptrons the same training images [15].

A number of types of learning could have taken place during the training. Novices would have acquired a familiarity with the appearance of a healthy femur and its bone tissues, and would have learnt to search the image for abnormal patterns. Some sensory-learning based enhancements in sensitivity may also have occurred [1]. Unlike Sowden, Davies and Roling [1] we did not attempt to isolate any one type of learning as this was not the focus of this study.

Models of medical image perception typically conceptualise the diagnostic process as involving a holistic first glimpse of the image, followed by a serial search process in locating any abnormal patterns and then reaching a decision on the significance of any abnormalities found [1, 24, 25]. Our training addressed only the first two components of these models. Accurate diagnosis of pathological features requires sophisticated understanding of disease pathogenesis for which the traditional, concept-based, analytical, rule-oriented approach remains important [10].

An open question is how generalizable are our findings. Because the images were selected at random, there is no reason to believe that our findings would not generalize to all other radiographs showing right hip fractures. In particular, no attempt was made to locate the fracture (if present) at a particular anatomical location. Similarly, no attempt was made to keep the orientation of the femur at a particular angle nor to keep the scale of the image precisely the same. Therefore, slight changes in these parameters are unlikely to affect our results. We did, however, ensure that all the shown images were of right hips, flipping left hip images to make them look as like right hip images, as needed. Thus, while we would expect most of the training to transfer to left hip images, it is possible that there might be a slight reduction in performance for left hip images, which in practice could be avoided by flipping these images so that they look like right hip images. Of more concern is to what degree our findings would extrapolate to diagnosing fractures of other bones. Presumably, there would be some transfer of learning, but the transfer would not be total. As the results of Experiment 1 and 2 showed, performance is maximised when observers are trained on exactly the sort of images that they are tested on and is reduced if the training set is not representative of the test set. As Experiment 2 demonstrated, constructing the training set to over-represent a subset of images reduces the effectiveness of the perceptual training. That said, some diseases are so rare that showing a few examples will misrepresent their naturally occurring frequency. While this cannot be avoided, our findings suggest that this should be minimised as far as possible.

In conclusion, our study demonstrates that perceptual training is likely to be a highly effective technique for training radiologists in at least some aspects of diagnosis and could be incorporated in the radiology curriculum. It may also be a particularly effective method of training non-radiologists (e.g. reporting radiographers and other physicians) in becoming proficient in a specific type of examination (e.g. chest x-rays for intensivists and emergency physicians, mammography for reporting radiographers). While some rare diseases can easily be over-represented by showing just a few examples, as far as possible the training set should accurately represent the sorts of images the radiologists would need to diagnose in practice at the frequencies at which these images would naturally occur. Repeating these images does not reduce the perceptual learning relative to showing an equivalent number of novel images. This reduces the number of images that are required for the perceptual training.

Supporting information

S1 File. Perceptual training dataset.
(XLSX)

S2 File. Appendix.
(DOCX)

Acknowledgments

This study was supported by a Royal Australian and New Zealand College of Radiology (RANZCR) research grant in 2015, and the University of Melbourne Engagement Grants Scheme 2015.

Author Contributions

Conceptualization: Weijia Chen, Frank Gaillard, Piers D. L. Howe.

Data curation: Weijia Chen.

Formal analysis: Weijia Chen.

Funding acquisition: Frank Gaillard, Piers D. L. Howe.

Investigation: Weijia Chen.

Methodology: Weijia Chen, Piers D. L. Howe.

Project administration: Weijia Chen, Piers D. L. Howe.

Resources: David HolcDorf, Mark W. McCusker, Frank Gaillard, Piers D. L. Howe.

Software: Weijia Chen, Piers D. L. Howe.

Supervision: Piers D. L. Howe.

Validation: Weijia Chen.

Visualization: Weijia Chen.

Writing – original draft: Weijia Chen.

Writing – review & editing: Weijia Chen, Frank Gaillard, Piers D. L. Howe.

References

1. Sowden PT, Davies IR, Roling P. Perceptual learning of the detection of features in X-ray images: a functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human perception and performance*. 2000; 26(1):379. PMID: [10696624](#)
2. Burton AM, Bruce V, Dench N. What's the difference between men and women? Evidence from facial measurement. *Perception*. 1993; 22(2):153–76. <https://doi.org/10.1068/p220153> PMID: [8474841](#)
3. Russell R. A sex difference in facial contrast and its exaggeration by cosmetics. *Perception*. 2009; 38(8):1211–9. <https://doi.org/10.1068/p6331> PMID: [19817153](#)
4. Komori M, Kawamura S, Ishihara S. Multiple mechanisms in the perception of face gender: Effect of sex-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37(3):626. <https://doi.org/10.1037/a0020369> PMID: [20822297](#)
5. Macrae CN, Martin D. A boy primed Sue: feature-based processing and person construal. *European Journal of Social Psychology*. 2007; 37(5):793–805.
6. O'Toole AJ, Deffenbacher KA, Valentin D, McKee K, Huff D, Abdi H. The perception of face gender: The role of stimulus structure in recognition and classification. *Memory & Cognition*. 1998; 26(1):146–60.
7. Kellman PJ. Perceptual Learning. In: Pashler H, Gallistel CR, editors. *Stevens' handbook of experimental psychology* (3rd edition). 3. New York: John Wiley & Sons; 2002.
8. Gibson EJ. *Principles of perceptual learning and development*. New York: Prentice-Hill; 1969.
9. Kellman PJ, Garrigan P. Perceptual learning and human expertise. *Physics of Life Reviews*. 2009; 6:53–84. <https://doi.org/10.1016/j.plrev.2008.12.001> PMID: [20416846](#)
10. Krasne S, Hillman JD, Kellman PJ, Drake TA. Applying perceptual and adaptive learning techniques for teaching introductory histopathology. *Journal of pathology informatics*. 2013; 4(1):34.
11. Xu B, Rourke L, Robinson JK, Tanaka JW. Training Melanoma Detection in Photographs Using the Perceptual Expertise Training Approach. *Applied Cognitive Psychology*. 2016; 30(5):750–6.
12. Young KG, Li RW, Levi DM, Klein SA. Interocular specificity in perceptual learning of a position discrimination task. *Investigative Ophthalmology and Vision Science*. 2005; 46:5631.
13. Ghose GM, Yang T, Maunsell JHR. Physiological correlates of perceptual learning in monkey V1 and V2. *Journal of Neurophysiology*. 2002; 87(4):1867–88. <https://doi.org/10.1152/jn.00690.2001> PMID: [11929908](#)

14. Petrov AA, Doshier BA, Lu ZL. The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*. 2005; 112(4):715–43. <https://doi.org/10.1037/0033-295X.112.4.715> PMID: 16262466
15. Haykin S. *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall; 1999.
16. LeCun Y. Efficient learning and second-order methods. A tutorial at NIPS 93, Denver. 1993.
17. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*. 2013; 310(20):2191. <https://doi.org/10.1001/jama.2013.281053> PMID: 24141714
18. Brainard DH, Vision S. The psychophysics toolbox. *Spatial vision*. 1997; 10:433–6.
19. Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*. 1997; 10(4):437–42. PMID: 9176953
20. Newell A, Rosenbloom PS. Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*. 1981; 1:1–55.
21. Enfron B, Tibshirani RJ. *An Introduction to the Bootstrap*. Cox DR, Hinkley DV, Reid N, Rubin DB, Silverman BW, editors: Taylor & Francis; 1993.
22. Green D, Swets J. *Signal detection theory and psychophysics*. New York: Wiley; 1966.
23. Biggs AT, Cain MS, Clark K, Darling EF, Mitroff SR. Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition*. 2013; 21(3):330–52.
24. Gale AG. *Human response to visual stimuli. The perception of visual information*: Springer; 1997. p. 127–47.
25. Drew T, Evans K, Võ ML-H, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*. 2013; 33(1):263–74. <https://doi.org/10.1148/rg.331125023> PMID: 23104971