Journal of Experimental Psychology:
Human Perception and Performance

© 2022 American Psychological Association
ISSN: 0096-1523

2022, Vol. 48, No. 12, 1336–1346
https://doi.org/10.1037/xhp0001057

Specific Versus Varied Practice in Perceptual Expertise Training

Samuel G. Robson[1], Jason M. Tangen[1], and Rachel A. Searston[2]
[1] School of Psychology, The University of Queensland
[2] School of Psychology, The University of Adelaide

We used a longitudinal randomized control experiment to compare the effect of specific practice (training on one form of a task) and varied practice (training on various forms of a task) on perceptual learning and transfer. Participants practiced a visual search task for 10 hours over 2 to 4 weeks. The specific practice group searched for features only in fingerprints during each session, whereas the varied practice group searched for features in five different image categories. Both groups were tested on a series of tasks at four time points: before training, midway through training, immediately after training ended, and 6 to 8 weeks later. The specific group improved more during training and demonstrated greater pre-post performance gains than the varied group on a visual search task with untrained fingerprint images. Both groups improved equally on a visual search task with an untrained image category, but only the specific group's performance dropped significantly when tested several weeks later. Finally, both groups improved equally on a series of untrained fingerprint tasks. Practice with respect to a single category (versus many) instills better near transfer, but category-specific and category-general visual search training appear equally effective for developing task-general expertise.

Public Significance Statement
The findings in this study suggest that training to find features for 10 hours—in a variety of images including fingerprints, aerial photographs, bark images, bone cancer images, retinas, and footwear impressions—can lead to robust and generalizable perceptual skill with fingerprints. This study also adds to previous work in demonstrating that practice with images from one category (fingerprints in this case) is more effective for improving performance on a trained task with the trained image category than practice with a wide breadth of image categories. Moreover, training across a wide variety of categories (rather than one) does not result in better performance with novel categories in the short term, but it may lead to more generalizable skill in the long term.

Keywords: expertise, fingerprints, skill transfer, visual search, varied practice

Supplemental materials: https://doi.org/10.1037/xhp0001057.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Depth and breadth during training can affect the degree to which a person improves at a task and how well they generalize their learning to new problems and examples. Does practicing a task in a multitude of forms improve learning and transfer to novel examples? In this study, we compared the effects of specific practice (training on only one form of the task) to varied practice (training on multiple forms of a task) on learning and transfer in perceptual domains. In particular, we explored how visual search training with images from a single category compares to training with images from several perceptual categories.

Perceptual experts in fields such as diagnostic medicine, sport, and forensic science regularly make consequential decisions. Becoming an expert in domains of this sort typically requires one to learn to efficiently identify and discriminate objects and patterns

This article was published Online First September 22, 2022.
Samuel G. Robson https://orcid.org/0000-0002-2777-9347
Jason M. Tangen https://orcid.org/0000-0002-0649-2566
Rachel A. Searston https://orcid.org/0000-0001-7295-8021
This research was supported by Grant LP170100086 from the Australian Research Council to Jason M. Tangen and Rachel A. Searston. We thank Gary Edmond, Kevin Eva, and Matthew Thompson for helping to secure funding. We thank Amy Cramb for assisting in data collection. We thank our partners and participants from across the country who made it possible to conduct this research.

Samuel G. Robson contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualization, validation, writing, and editing. Jason M. Tangen contributed to conceptualization, methodology, software, validation, editing, supervision, resources, and funding acquisition. Rachel A. Searston contributed to methodology, software, visualization, funding acquisition, supervision, and editing. All authors have read and approved the final article.
Correspondence concerning this article should be addressed to Samuel G. Robson, School of Psychology, The University of Queensland, Sir Fred Schonell Drive, St. Lucia 4072, Queensland, Australia. Email: s.robson@uq.edu.au

(Shen et al., 2014). Almost by definition, expertise is domain specific—we would not expect a chess grand master to be exceptional at identifying bird species. A vast literature on perceptual expertise supports the idea that skills are typically specific to the stimuli that one is exposed to during training (e.g., Chase & Simon, 1973; Curby et al., 2009; Diamond & Carey, 1986). For example, radiologists can locate abnormalities in medical images quicker than novices, but they perform like novices if the medical images are inverted or when given generic search puzzles like "Where's Waldo" (Chin et al., 2018; Nodine & Krupinski, 1998). Specificity of this sort can apply to low-level perceptual dimensions such as orientation and motion (Ahissar & Hochstein, 1997; Ball & Sekuler, 1987; Fahle, 2005; Fiorentini & Berardi, 1980) and high-level perceptual categories (Bukach et al., 2010; Chase & Simon, 1973; Diamond & Carey, 1986; Ericsson, 2017; Ericsson & Lehmann, 1996; Gauthier et al., 1998). In line with stimulus specificity and domain specificity, training with a single image category, rather than a variety of image categories, ought to result in better performance on tasks that involve the trained image category.

Generalizing one's skills to new categories, however, is a separate problem. Studies from cognitive, motor, and verbal learning literatures have shown that varied practice results in better transfer to novel variants of a task and sometimes even better performance on the trained task variant (Goode et al., 2008; Heitman et al., 2005; Kerr & Booth, 1978; Landin et al., 1993; Roller et al., 2001; Shea & Kohl, 1990; Vakil & Heled, 2016; Willey & Liu, 2018). Greater breadth during training may promote more general problem-solving strategies and enable more elaborative processing, allowing for a wider variety of associations between stimuli and task parameters (see Craik & Tulving, 1975; Goode et al., 2008; Schmidt, 1975; Shea & Kohl, 1990). Similar mechanisms may operate in perceptual learning contexts—training with multiple categories rather than one might result in better performance with novel categories.

In the present study, we explored the merits of depth versus breadth of practice. We compared the effects of specific versus varied practice on transfer to novel instances, novel image categories, and novel tasks in a way that is analogous to comparing domain-specific to domain-general practice. As we have discussed, domain-specific practice tends to instill exceptional but limited perceptual skill, whereas more varied practice may instill less exceptional but more generalizable perceptual skill. In this experiment, we investigated these ideas using fingerprint identification as a case domain.

Fingerprint Identification and Visual Search

Human examiners are employed by police and security agencies to decide whether a fingerprint left at a crime scene originated from the same finger as another print collected from a suspect. Oftentimes the primary objective of these examiners is to judge whether the two different fingerprint impressions match. Fingerprint examiners typically receive years of specialized training, but fingerprints as a category are novel to almost everyone else. Indeed, these examiners can match prints far better than novices (Tangen et al., 2011; Thompson, Tangen, & McCarthy, 2014), and they have superior performance on a range of other fingerprint-specific tasks by virtue of their training (Busey & Vanderkolk, 2005; Robson et al., 2021; Searston & Tangen, 2017a, 2017b; Thompson & Tangen, 2014; Thompson, Tangen, & Searston, 2014; Vogelsang et al., 2017). Expertise with fingerprints is therefore domain specific and trainable. Fingerprints thus serve as an ideal category for comparing the effects of varied and specific practice on skill acquisition and generalizability.

In prior studies, researchers have had success training perceptual skill in laboratory settings often using tasks where participants are exposed to many category instances that must be sorted correctly (Gauthier et al., 1998; Searston & Tangen, 2017c; Scott et al., 2008; Tanaka et al., 2005; Wong et al., 2009). However, developing perceptual expertise relies heavily on learning to search for, attend to, and extract useful or diagnostic information (Chua et al., 2014, 2015; Jiang & Chun, 2001; Kellman & Garrigan, 2009). Restricting or cuing participants to diagnostic information can improve decision-making accuracy in tasks such as classifying fish (Baruch et al., 2014), identifying aircraft (Dror et al., 2008), and matching unfamiliar faces (Towler et al., 2021). Visual search training can also benefit performance in radiology and baggage screening tasks (Auffermann et al., 2018; Nakashima et al., 2013; Schuster et al., 2013). Developing perceptual skill with fingerprints may similarly rely on attention and visual search.

The process of examining prints has been described as a thorough analysis of minutiae where one carefully searches for corresponding and discordant features on the two comparison prints (SWGFAST, 2012; Vanderkolk, 2011). Compared to novices, fingerprint examiners attend to more constrained sets of features (Roads et al., 2016; Robson et al., 2020). Expert examiners also more efficiently localize useful features in typical, intact fingerprints but not scrambled fingerprints (Robson et al., 2021). Searching for features is therefore an important process in deciding whether two prints match. This research also highlights how expert visual search ability tends to be specific to the image category or domain an expert is familiar with. A visual search paradigm is therefore ideal for investigating how different practice regimes might affect performance with respect to fingerprints.

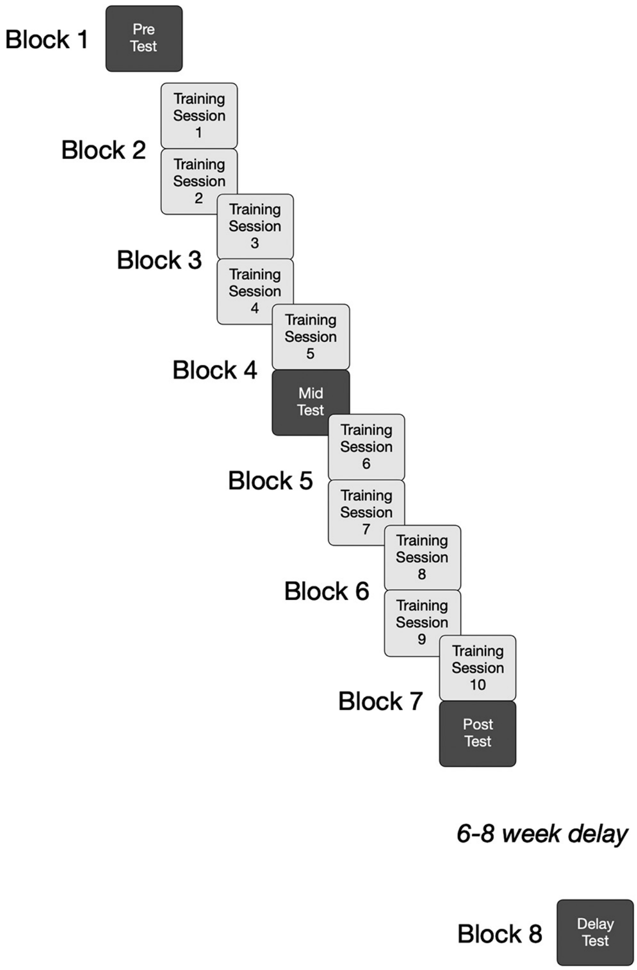
The Current Study

In the current study, participants with no formal experience in fingerprint examination were randomly allocated to one of two visual search training interventions, each taking place over 10 1-hour sessions (as illustrated in Figure 1). Four testing sessions were also administered throughout these interventions. The training and testing took place over eight "blocks" administered on separate days over a period of several weeks.

In the training sessions, participants were instructed to localize a small fragment of visual information within a larger image array in a task that we refer to as "Find-the-Fragment." We have demonstrated in a prior study that experts outperform novices on a task of this sort (Robson et al., 2021). In the current study, one group trained on this Find-the-Fragment task only with fingerprint images (specific practice), whereas the other trained with five different image categories (varied practice). Participants also completed a series of tests at four points: immediately prior to the first practice session (pretest), after 5 hours of practice (midtest), after 10 hours of practice (posttest), and several weeks after the final practice session (delay test). Included in these testing sessions were two tests of visual search ability—one with a trained image category and one with an untrained image category—as well as a

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 1
Time Course of the Training and Testing Sessions



Note. Participants visited the lab on eight occasions, completing 10 training sessions and four testing sessions.

general measure of fingerprint expertise. This design allowed us to measure the degree to which each group improves (a) across the training sessions, (b) on new instances of the trained task with the trained image category, (c) on a version of the trained task with an untrained category, and (d) on untrained tasks with the trained image category (i.e., fingerprints), with the latter also serving as a measure of expertise in fingerprint identification.

Method

Transparency and Openness

This study was approved by the University of Queensland Human Research Ethics Committee. The rationale, methods, sensitivity analyses, exclusion criteria, hypotheses, and planned analyses were preregistered on the Open Science Framework, and the project can be found at https://osf.io/96d2w/. Additionally, all deidentified data and analysis scripts can be found in the "Analyses" component of this project. The experimental code is also available in the "Materials" component, excluding the images used as some data sets contain sensitive or copyrighted material.

Participants

A total of 43 participants completed the training program. Of these, 21 were allocated to the specific practice condition (17 female, four male; *Mdn* age = 24.0), and 22 were allocated to the varied practice condition (14 female, eight male; *Mdn* age = 23.5). Rather than aiming to increase the chance of detecting an effect by gathering data from many dozens of participants, we instead focused on increasing the duration of the training intervention so as to increase the size of the effect we aimed to detect. Given effect sizes found in prior perceptual training studies and studies comparing specific and varied practice (e.g., Gauthier et al., 1998; Kerr & Booth, 1978; Roads et al., 2016; Tanaka et al., 2005; Willey & Liu, 2018), and the sheer length of our training intervention, we conducted power analyses to detect what we consider a conservative effect (Cohen's $d = .45$). A total of 40 participants (minimum 20 replicates) provided adequate power ($> .8$) to detect an effect of this magnitude across all planned analyses.

We recruited 73 participants through the University of Queensland's paid participant pool to complete the pretest, and from this pool we asked participants if they wished to take part in further training. Of these participants, 15 were unavailable or did not wish to take part in further training, and another 15 did not meet inclusion criteria (13 performed below 25% accuracy at pretest and two did not complete all training sessions). These 30 participants were excluded from the analyses. Additionally, 26 of the participants returned to complete a delay test administered 6 to 8 weeks after the last training session. All participants were paid at a rate equivalent to 20 Australian dollars per hour either in gift cards or cash. We also incorporated a scoring system in the training sessions where participants could earn an additional 5 dollars each session if they surpassed their highest score from their prior sessions.

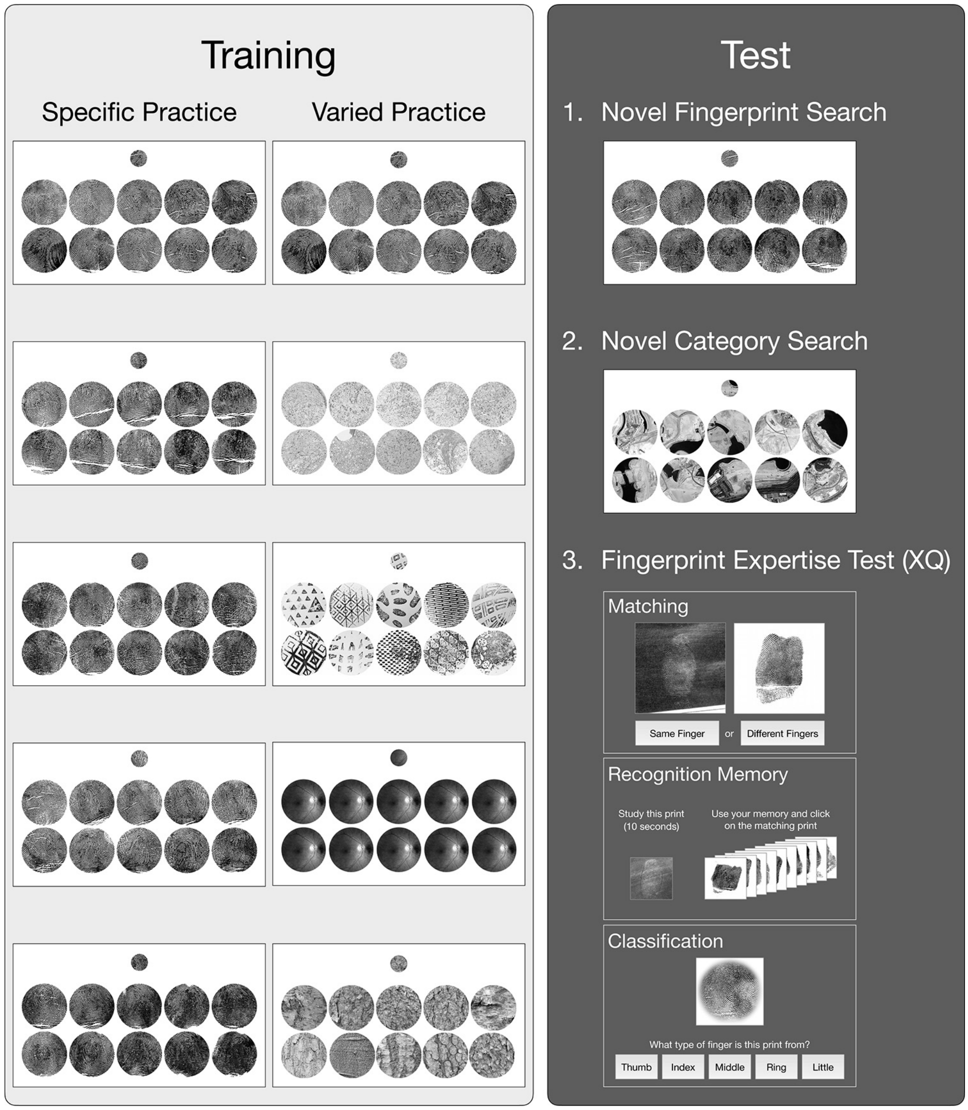
Design

We employed a longitudinal, between-subjects (specific vs. varied) randomized experimental design. Participants visited the lab on eight occasions (or blocks) over 2 to 4 weeks, completing 10 hours of training and four sessions of testing. The number of practice hours were based on similar prior studies (e.g., Gauthier et al., 1998). A summary of the training task and tests is displayed in Figure 2.

In each training session, participants completed as many find-the-fragment trials as possible in the allocated time. During each trial, participants were instructed to search for a small circular "fragment" of visual detail in an array of 10 images from a particular category. Those in the specific practice condition searched only for fingerprint fragments in arrays of fingerprints. Those allocated to the varied practice condition completed five different versions of the Find-the-Fragment task, including the fingerprint version and four of five other category versions (aerial photographs, bark, retinas, osteosarcoma histology images, and shoeprints). These four categories were chosen at random for each participant in the varied condition for their training sessions, with the fifth reserved to test performance on an untrained image category (see

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 2
Example Training and Testing Sessions



Note. Participants were randomly allocated to one of two training conditions (specific vs. varied). During training, the specific group searched for features in only one category of images (fingerprints), whereas the varied group searched for features in a variety of image categories. Each group completed three different tests at each testing point: a visual search test with untrained fingerprint images, a visual search test with an untrained image category, and the XQ—a multitask test of perceptual expertise with fingerprints, which served as a test of transfer to untrained fingerprint tasks. These example items are for display purposes and may not have been presented in the experiment.

Procedure). The order of the trials, and the images presented, were randomized for each participant.

Participants completed three tests at four time points: before training commenced (pretest), midway through training (midtest), at the end of training (posttest), and 6 to 8 weeks after training (delay test). The first of these tests was a fingerprint Find-the-Fragment task with untrained images. The second test was a find-fragment task with an image category that the participant had not encountered during training. The final test was a measure of fingerprint expertise called the "XQ" (Searston et al., 2021), which consisted of three untrained subtasks: a fingerprint matching task, a fingerprint recognition memory task, and a hand and finger classification task. The images and image order across all tests were unique to each participant.

Material

In all the training sessions, and in two of the tests, participants engaged in some variation of the Find-the-Fragment task. In this task, participants were presented with an array of 10 images of a particular category as well as a small circular fragment presented

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

above the array, which had been cropped from one of the 10 images. Participants were instructed to click on the location of the fragment in the image array.

To create these arrays, we sourced thousands of images from large databases: 2,620 fingerprints from the National Institute of Technology's Special database 300 (Fiumara et al., 2018); 1,021 aerial photographs from the Maritime Satellite Imagery database (Gallego et al., 2018); 659 bark images from the BarkNet 1.0 database (Carpentier et al., 2018); 557 images from the osteosarcoma database in the Cancer Imaging Archive (Clark et al., 2013; Leavey et al., 2019); 1,058 retina images from the Messidor 2 database (Abràmoff et al., 2013; Decencière et al., 2014); and 591 images from the Footwear Impression database (Kortylewski et al., 2015). We cropped and rescaled every image to a 500 × 500 circle and converted them to grayscale. Next, we calculated the cosine similarity between the pixels in each image and in every other image within the same category. Using these similarity ratings, images were sorted into groups of 10 highly similar instances. We then removed any image groups that appeared too easy or lacked clarity. Finally, we created 20 different 2 × 5 image grids for each image group (1,390 × 550 pixels; 10 pixels between each image), randomizing the position of the images within the grid each time. A total of 11,340 arrays were generated (5,240 fingerprint arrays; 1,700 aerial arrays; 1,000 bark arrays; 1,000 osteosarcoma histology arrays; 1,300 retina arrays; and 1,100 shoeprint arrays).

To generate the fragment images, we selected a random location in one of the images in each array and copied out a circle of 100 × 100 pixels. Two fragments were extracted from every single image. In total, 11,340 fragments were generated, one for each array.

Procedure

All tasks and instructions were presented on a 13-in. MacBook laptop screen with over-ear headphones. LiveCode (Version 9.0) was used to present stimuli and record responses. At the beginning of each training and testing session, participants provided demographic information.

Training Sessions

In the very first training session, participants watched video instructions explaining the Find-the-Fragment task. At the beginning of each trial, a fragment and array appeared on the screen. A timer beginning at 30 was also displayed at the screen's top left corner. The fragment and array images remained on the screen until the participant clicked on the array. Meanwhile, the timer counter dropped by 1 every second. If participants clicked on the fragment's correct location, a light tone sounded, and a green tick appeared. If participants clicked directly on the fragment, or within 5 pixels of the fragment, their response was coded as correct. If correct, the number remaining on the timer was added to the participant's score for the session, which was displayed at the screen's top right corner. If participants were incorrect, a dull tone sounded, a red cross appeared, and nothing was added to the participant's score. In both cases, a blue highlight of the fragment's correct location was displayed. Feedback remained on the screen for 2,000 ms before the next set of images appeared along with a new timer.

During each session, participants engaged in the Find-the-Fragment task for 50 min or until they completed 480 trials. If a participant surpassed their highest score from their previous sessions, they received additional payment, which they were reminded of before every session. The participant's highest score from their previous sessions was displayed above their current session's score.

Response time and accuracy (correct or incorrect fragment selection) were recorded on every trial. We removed all trials with response times quicker than 500 ms and slower than 60 s and then combined accuracy and response time into a single integrated speed-accuracy measure known as the balanced integration score (BIS; Liesefeld et al., 2015). We computed the BIS separately for the training data and for each of the visual search tests (see below). In each case, we standardized the proportion correct across all trials, and the response times on the correct trials, for each participant at each time point before then subtracting one standardized score from the other. A BIS of 0 is the average score across all cells, and the standard deviation is equal to 1. Positive values indicate scores above the mean, and negative scores indicate performance below the mean. Someone with a higher score is relatively more accurate and quicker on task, whereas someone with a lower score is relatively inaccurate and slower to respond. The BIS was our primary measure of interest because it accounts for individual differences in speed-accuracy trade-offs (see Liesefeld & Janczyk, 2019).

Testing Sessions

There were four testing sessions in total: a pretest, midtest, posttest, and delay test (administered 6 to 8 weeks after the final training session), each consisting of the same three tests. Every participant was presented with a unique set of test images that differed from the images they saw during training. These images were presented in the same order at each testing point, but feedback was never given. Instructions were presented prior to each task.

Fingerprint Find-the-Fragment Test. To measure how well participants' search skills transferred to untrained images on the trained visual search task, we tested them on a fingerprint version of the Find-the-Fragment task with images they had not seen during training. Performance was measured using the BIS.

Novel Category Find-the-Fragment Test. To measure how well participants' search skills transferred to untrained image categories, we tested them on the Find-the-Fragment task with images from the category that they had not encountered during training. For example, if a participant trained on fingerprints, aerial photographs, bark images, osteosarcoma images, and shoeprints, they were tested on retina images. The untrained category was randomized for each participant, and performance was measured using the BIS.

Fingerprint Expertise Test (XQ). To measure whether visual search training can improve general fingerprint expertise, and by extension whether visual search training transfers to untrained tasks with the trained image category, participants were tested using the XQ (Searston et al., 2021). This test was created by inviting 44 professional fingerprint examiners and 44 matched, untrained novices to complete 10 tasks designed to capture a variety of domain-specific perceptual skills required for fingerprint

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

identification. From there, the most optimal combination of tasks for discriminating between experts and novices in the shortest time frame was selected. A subset of three tasks was identified through this process to serve as a brief, general test of perceptual expertise in fingerprint examination. The three subtests included a fingerprint matching task, a fingerprint recognition memory task, and a hand and finger classification task. All three subtests of the XQ take approximately 12 min to complete together. The fingerprint images used were randomly sampled from a set of 3,000 prints.

The matching component of the XQ consists of 12 trials. In this task, two prints are presented side by side, and participants indicate whether the print pair originated from the same finger or different fingers. The recognition memory component of the XQ consists of 12 trials. On each trial, a crime scene fingerprint is presented for 10 s, which participants are prompted to study. The print then disappears, and participants search through a set of 10 fully rolled prints one by one. They then select the one that they think matches the studied print. The classification component of the XQ consists of 10 trials. A single print is presented, and participants must indicate which hand (left or right) and finger type (thumb, index, middle, ring, and little) the print was sourced from. Proportion correct for hand classification and finger classification are averaged. Overall XQ scores are calculated by averaging participants' proportion correct on each of the three subtests.

Hypotheses

Those in the specific practice group encountered only fingerprint images during their training, whereas those in the varied practice group encountered a range of different image categories. The specific group was therefore exposed to a more narrowly defined visual task than the varied group by virtue of practicing only with fingerprint images across all 10 training sessions. We therefore hypothesized that the specific practice group would improve more than the varied practice group across the 10 training sessions (Hypothesis 1).

The varied practice group was exposed to many more fingerprint images and features compared to the specific practice group throughout training. In line with stimulus-specific and domain-specific accounts of skill acquisition, the specific group ought to have become more attuned than the varied group to the variation and structural regularities within fingerprints. We hypothesized that the specific group would therefore improve more than the varied group from pretest to posttest on the Find-the-Fragment test with novel fingerprint images (Hypothesis 2).

Across a range of tasks, varied practice results in skills that better generalize to untrained task variants, when compared to specific practice (e.g., Goode et al., 2008; Kerr & Booth, 1978; Vakil & Heled, 2016; Willey & Liu, 2018). One explanation for this effect is that varied practice promotes more general problem-solving strategies and elaborative processing (see Craik & Tulving, 1975; Schmidt, 1975; Shea & Kohl, 1990). Applying these findings to a visual search task, such as our Find-the-Fragment test, those who practice with a variety of image categories should develop generalizable visual search strategies that are not anchored to any one class of images—as demonstrated by improved performance with novel image categories. In the current study, we therefore hypothesized that the varied practice group would improve more than the specific practice group from pretest to posttest on the Find-the-Fragment test with the untrained image category (Hypothesis 3).

Any advantage that varied practice may elicit, however, ought to apply only to the dimension that varied during training, which in the current study was the image category. Practice with a variety of image categories might help learners to generalize their search skill to novel categories, but not necessarily to novel tasks. A domain-specific account, however, might predict that more time practicing with fingerprints ought to result in superior performance with fingerprint-related tasks more generally. We therefore expected the specific practice group to improve more than the varied practice group on our general measure of fingerprint expertise (the XQ), which consists of three untrained tasks (Hypothesis 4). We also expected performance gains on the XQ to be driven by increases in matching and recognition memory performance, rather than increases in hand and finger classification accuracy, because only the former two abilities seem related to visual search.

Finally, we conducted exploratory analyses to assess the degree to which participants retained their perceptual skills 6 to 8 weeks after the final training session. These analyses can inform whether specific and varied visual search training induces robust and generalizable improvements in performance on trained and untrained tasks and image categories or whether the improvements reflect more immediate, superficial performance gains (see Soderstrom & Bjork, 2015).

Results

We conducted a series of linear mixed models to compare how much each group improved throughout training and on each of the tests. For each model, we specified the fixed effects to be condition (specific, varied), training/testing session, and the interaction between these factors. Participant was included as the random effect. Follow-up tests included Bonferroni corrections. The results are displayed in Figure 3. Additional analyses, including data on response time and proportion correct, can be found in the online supplemental materials.

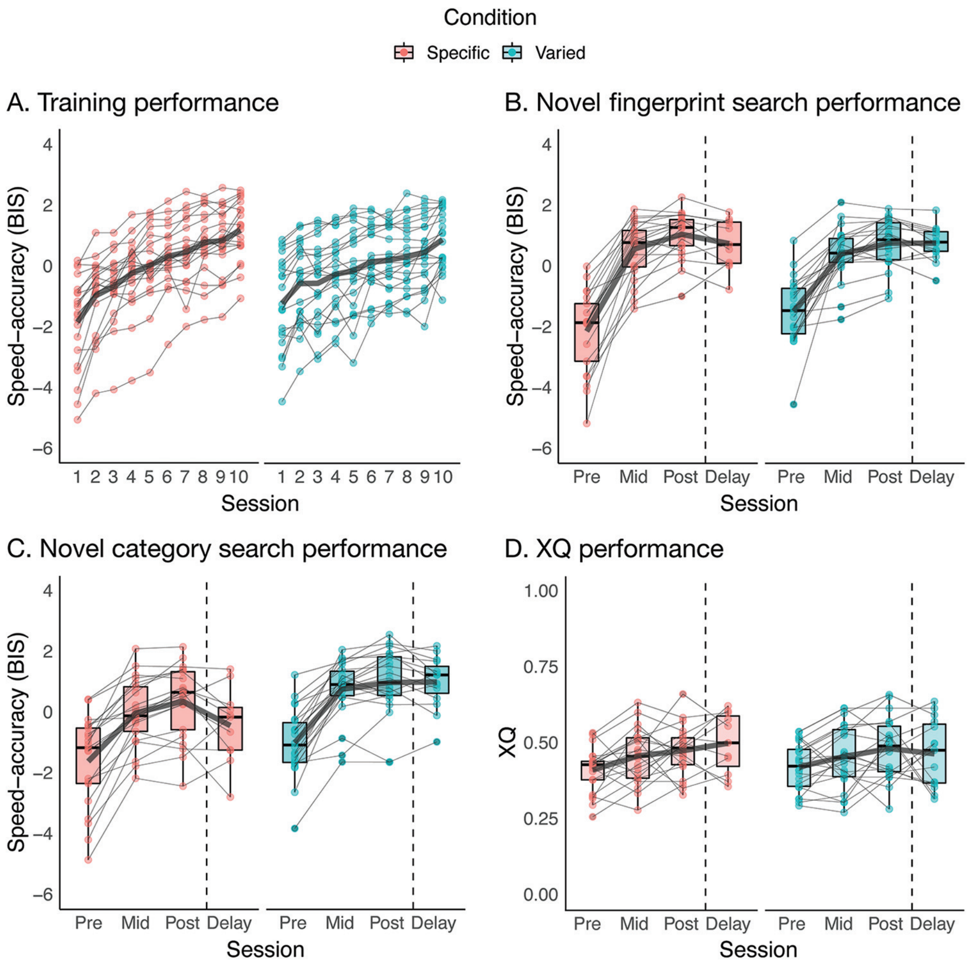
Training Performance

To analyze how much each group improved on the trained task, we assessed participants' speed-accuracy (measured via BIS) over the 10 training sessions. On average, participants completed a total of 3,335 Find-the-Fragment trials over the 10 sessions. The linear mixed model explained 88.9% of the variance in performance. The fixed effects explained 26.3% (conditional intraclass correlation coefficient [ICC] = .626). The model revealed a significant interaction between condition and training session, $F(1, 385) = 40.85$, $p < .001$. We followed up this interaction with simple slopes analyses, which revealed that both the specific practice group ($b = .30$, $SE = .01$, $p < .001$) and varied practice group ($b = .20$, $SE = .01$, $p < .001$) improved over the 10 training sessions, but the slope for the specific practice group was steeper. Together, these analyses suggest that the specific practice group improved to a greater extent than the varied practice group across the training sessions.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 3

Training and Transfer Test Performance Over Time



Note. The results for both the specific (red, left) and varied practice (blue, right) conditions across the training sessions (A) and on each of the transfer tests (B, C, D). Gray lines connect participants' scores across the sessions, and the larger gray lines connect the cell means. The dashed, vertical lines denote the 6- to 8-week interval between the posttest and delay test. BIS = balanced integration score. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

An additional between-groups *t* test revealed that the two groups did not differ in performance in the first session, $t(41) = 1.56$, $p = .126$. Fingerprints were also the third most difficult category according to the speed-accuracy scores during training. The two training interventions therefore appear to be of similar difficulty, and the significant interaction is unlikely to be the result of a ceiling effect.

Fingerprint Search Test: Transfer to Novel Images From the Trained Category

We tested transfer to untrained images from the trained category by comparing speed-accuracy on the fingerprint Find-the-Fragment test at pretest, midtest, and posttest. The linear mixed model explained 80.6% of performance variance. Fixed effects explained 59.7% of variance in performance (conditional ICC = .209). There was no main effect of condition, $F(1, 41) = .08$, $p = .782$, but there was a main effect of testing session, $F(2, 82) = 192.56$, $p < .001$. Following up this main effect revealed that participants improved overall from pretest to midtest, $t(82) = 15.39$, $p < .001$, and from midtest to posttest, $t(82) = 2.84$, $p = .017$. The effect size for the pre-post improvement was very large ($d = 3.93$).

There was also a significant interaction between condition and testing session, $F(2, 82) = 6.01$, $p = .004$. Follow-up simple comparisons revealed that the specific group improved significantly from pretest to midtest, $t(82) = 12.69$, $p < .001$, but not from midtest to posttest, $t(82) = 2.29$, $p = .073$. Similarly, the varied practice group improved from pretest to midtest, $t(82) = 9.04$, $p < .001$, but not significantly between midtest and posttest, $t(82) = 1.72$, $p = .267$. Overall, the pre-post improvement was larger for the specific group ($d = 4.62$) than the varied group ($d = 3.25$). Thus, the specific group improved to a greater extent than the varied group on novel category variants of the fingerprint Find-the-Fragment test. A between-groups *t* test indicated that the two groups did not

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

differ statistically at pretest, $t(39) = 1.74$, $p = .089$, suggesting that a ceiling effect is an unlikely explanation for this interaction.

We ran exploratory analyses to test whether performance on the fingerprint Find-the-Fragment test was maintained between the posttest and delay test administered 6 to 8 weeks later. The linear mixed model explained 78.8% of performance variance. Fixed effects explained 5.9% of the variance (conditional ICC = .729). We found a significant main effect of test session, $F(1, 26) = 4.60$, $p = .042$, but no main effect of condition or an interaction (ps $>$ .05). Participants therefore showed a statistically significant decline in performance between the posttest and delay test.

Novel Category Search Test: Transfer to Novel Image Categories

We tested transfer to untrained image categories by comparing participants' speed-accuracy on the Find-the-Fragment test with an untrained image category at pretest, midtest, and posttest. A between-groups t test revealed that the performance of the two groups did not differ at pretest, $t(37) = 1.53$, $p = .136$. The linear mixed model explained 87.3% of performance variance. Fixed effects explained 39.0% of variance in performance (conditional ICC = .483). Contrary to our expectations, there was no significant interaction between condition and testing session, $F(2, 82) = .34$, $p = .715$. However, there was a main effect of condition, $F(1, 41) = 4.10$, $p = .0496$, indicating that the varied group outperformed the specific group overall. There was also a main effect of testing session, $F(2, 82) = 169.89$, $p < .001$. A follow-up of this main effect revealed that performance increased between pretest and midtest, $t(82) = 14.51$, $p < .001$, and between midtest and posttest, $t(82) = 2.60$, $p = .033$. The effect size for the overall performance increase from pretest to posttest was very large (d = 3.69).

We ran exploratory analyses to test whether the performance gains on the Find-the-Fragment test with the untrained category were maintained between the posttest and the delay test administered several weeks later. The linear mixed model explained 76.1% of performance variance, with fixed effects explaining 11.6% of variance in performance (conditional ICC = .644). Overall, there was a statistically significant decrease in performance between posttest and delay test, $F(1, 27) = 5.03$, $p = .033$. A main effect of condition, $F(1, 43) = 5.59$, $p = .023$, also indicated that the varied group outperformed the specific group across the two sessions. The interaction was nonsignificant, $F(1, 27) = 1.02$, $p = .322$, but follow-up tests revealed that the specific group's performance declined significantly between posttest and delay test, $t(28) = 2.14$, $p = .042$ ($d = -.89$), whereas the varied group's performance did not differ statistically between posttest and delay test, $t(27) = .94$, $p = .357$ ($d = -.34$). Thus, some evidence indicates that the varied group was able to retain the same level of skill on the search test with novel categories several weeks after training, whereas the specific group's performance dropped significantly.

Test of Fingerprint Expertise: Transfer to Novel Tasks

Last, we assessed transfer to untrained tasks by comparing performance at pretest, midtest, and posttest on a general test of fingerprint expertise (the XQ), which consisted of three untrained fingerprint tasks. The dependent variable was the average proportion correct across the three tasks. The two groups did not differ at pretest, $t(41) = .56$, $p = .558$. The linear mixed model explained 54.5% of the variance in performance. Fixed effects explained 8.0% of the variance in performance (conditional ICC = .465).

Contrary to our expectations, there was a nonsignificant interaction between condition and testing session, $F(2, 82) = .13$, $p = .876$. There was also a nonsignificant main effect of condition, $F(1, 41) = .11$, $p = .742$. However, there was a significant main effect of testing session, $F(2, 82) = 10.93$, $p < .001$, indicating that overall performance on the measure of fingerprint expertise increased from pretest to posttest. Following up this main effect revealed that participants' scores improved from pretest to midtest, $t(82) = 2.96$, $p = .014$ ($d = .64$), but not significantly between midtest and posttest, $t(82) = 1.65$, $p = .308$ ($d = .36$). The overall effect size for the increase in XQ scores between pretest and posttest performance was large ($d = 1.00$). Additionally, the gains on the XQ were largely driven by improved matching ability (50.0% at pretest to 58.7% posttest) and recognition memory ability (37.0% at pretest to 44.2% posttest), whereas hand and finger nomination ability improved less by comparison (38.0% at pretest to 40.8% posttest). See the online supplemental materials for further details.

Finally, we tested whether participants' XQ scores changed between the posttest and delay test. The linear mixed model explained 46.2% of performance variance. Fixed effects explained 7.7% of variance in performance (conditional ICC = .455). No significant main effects or interactions were found (ps $>$ .05). Thus, neither group's performance changed significantly between the posttest and when tested several weeks later.

Discussion

The purpose of this study was to investigate the effect of specific versus varied practice on the acquisition and generalizability of perceptual skills, and we used fingerprint identification and visual search training as means to investigate our research questions. Participants trained on a visual search task for 10 1-hour sessions during which they located small fragments of visual detail in image arrays. Some practiced with fingerprint images only (specific practice), whereas others trained with images from five image categories (varied practice). We tested performance on three different tests at four time points: before training, midway through training, immediately after training, and 6 to 8 weeks after training.

We found that performance during the training sessions improved over time, but those who trained only with fingerprint images improved more than those who trained with images from a variety of image categories, as hypothesized (Hypothesis 1). This finding is unsurprising because those who engaged in specific practice needed to learn only about the visual structure and features of one image category, whereas those who underwent varied practice needed to become attuned to the visual structure and features of five image categories within the same period. Similarly, by virtue of regularly switching from one category to another in the varied practice condition, contextual interference (Battig, 1972) may have suppressed performance during training.

On the visual search test with untrained fingerprint images, we again found that those who trained only with fingerprints improved more than those who trained with a variety of image categories, as

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

expected (Hypothesis 2). Those who practiced the visual search task only with fingerprints spent 5 times longer studying fingerprints compared to those who practiced the task with a variety of image categories. The specific group likely developed a greater appreciation for fingerprint features and the locations where they tend to appear, resulting in better visual search performance even with novel prints. This finding aligns with a domain-specific account of perceptual skill (see Bukach et al., 2010; Chase & Simon, 1973; Diamond & Carey, 1986; Ericsson, 2017; Ericsson & Lehmann, 1996; Gauthier et al., 1998) where more time spent practicing in a particular domain, or with a particular image category, engenders greater skill with that specific image class.

Counter to our predictions, the varied practice group did not improve more than the specific practice group from pretest to posttest on the visual search test with the untrained image category (Hypothesis 3). Instead, those who practiced the visual search task with images from multiple categories, and those who practiced only with fingerprints, improved on this test to the same degree. Practicing with a variety of image categories did not result in more general search skills. However, this is not to say that varied practice cannot provide transfer benefits in other contexts and tasks (see, Goode et al., 2008; Roller et al., 2001; Vakil & Heled, 2016; Willey & Liu, 2018). It is also possible that the varied group could have developed more general search strategies had we greatly increased the number of categories they practiced with throughout training.

Exploratory analyses, nonetheless, provided some indication that participants who practiced with several image categories (rather than one) better retained their skill on the search test with novel categories when tested several weeks after training. Whereas the varied group's performance remained roughly the same between the posttest and delay test, the specific group's performance declined significantly. It may be that practicing with varied image categories results in better generalizability, but the effect may not be immediately apparent. Performance that arises immediately as a result of practice can be distinguished from learning, which is a relatively permanent acquisition of a skill (see Soderstrom & Bjork, 2015, for review). Learning can sometimes occur without any immediate gains in performance, and performance can improve without any long-term skill acquisition. The visual search performance of both groups improved throughout training, but perhaps the varied group learned this skill more effectively and retained it better in the long term. This interpretation, however, should be entertained only tentatively as the associated interaction was nonsignificant and the sample size was relatively small. Future studies may shed more light on the distinction between performance and learning in similar contexts.

Participant performance on a general measure of fingerprint expertise, which consisted of three untrained tasks, also improved from pre- to posttest by an average of 1 standard deviation. No statistical difference between scores at the posttest and delay test also suggests that this improvement was relatively enduring. Counter to our prediction that the specific group would improve more than the varied group on these tasks (Hypothesis 4), however, those who trained with fingerprints only improved to the same extent as those who trained with multiple image categories. Greater exposure and sensitivity to the specific visual structure of fingerprints is therefore unlikely to be responsible for the increases in fingerprint expertise scores. Instead, localizing complex visual information of any sort for several hours may underlie these improvements. Another explanation is that exposure to fingerprints in the varied practice condition, although reduced compared to the specific condition, was sufficient to reach a similar training saturation point. A condition where learners train with several categories, but never fingerprints, could potentially tease apart these explanations.

The aim of this experiment was to compare the relative merits of specific practice and varied practice. However, we cannot be sure that the visual search training intervention was the cause of the overall performance gains on each of our tests. An additional control group that received no training would be required to determine the absolute performance benefits of the training intervention. However, there are several reasons why the intervention is the likely cause for the improvements on the XQ. A validation of this test found that test-retest performance in the absence of training was stable (Searston et al., 2021). Participants also never received feedback at any point during the testing sessions. Moreover, performance gains were largely driven by improved fingerprint matching and recognition memory ability, but not hand and finger nomination ability—only the former two abilities appear to involve some degree of visual search.

There are of course many skills that underlie perceptual expertise with fingerprints (see Growns & Martire, 2020). Practice across a suite of tasks will likely aid the development of fingerprint expertise to a greater degree than just one task, but visual search practice may be worth integrating into fingerprint examiner training. We also noticed a great degree of variation between participants in this experiment—a better understanding of individual differences in perceptual ability, motivation, and response to training warrants further investigation.

Conclusions

In this study, we set out to compare specific practice and varied practice on learning and transfer using a visual search paradigm. We found that practice with respect to one image category (versus many image categories) results in better transfer to novel instances of a trained image category. We did not find evidence that practice with many image categories results in better immediate transfer to novel categories, but some evidence indicated a benefit in the long term. General visual search training may also be one method for aiding the development of expertise in real-world perceptual domains.

References

Abràmoff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D. C., Quellec, G., & Niemeijer, M. (2013). Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmology*, *131*(3), 351–357. https://doi.org/10.1001/jamaophthalmol.2013.1743

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. https://doi.org/10.1038/387401a0

Auffermann, W. F., Krupinski, E. A., & Tridandapani, S. (2018). Search pattern training for evaluation of central venous catheter positioning on chest radiographs. *Journal of Medical Imaging*, *5*(3), Article 031407. https://doi.org/10.1117/1.JMI.5.3.031407

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*(6), 953–965. https://doi.org/10.1016/0042-6989(87)90011-3

Baruch, O., Kimchi, R., & Goldsmith, M. (2014). Attention to distinguishing features in object recognition. *Visual Cognition*, *22*(9–10), 1184–1215. https://doi.org/10.1080/13506285.2014.987860

Battig, W. F. (1972). Intratask interference as a source of facilitation in transfer and retention. In R. F. Thompson & J. F. Voss (Eds.), *Topics in learning and performance* (pp. 131–159). Academic Press.

Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Attention, Perception, & Psychophysics*, *72*(7), 1865–1874. https://doi.org/10.3758/APP.72.7.1865

Busey, T. A., & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research*, *45*(4), 431–448. https://doi.org/10.1016/j.visres.2004.08.021

Carpentier, M., Giguere, P., & Gaudreault, J. (2018). *Tree species identification from bark images using convolutional neural networks*. arXiv. https://doi.org/10.48550/arXiv.1803.00949

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Chin, M. D., Evans, K. K., Wolfe, J. M., Bowen, J., & Tanaka, J. W. (2018). Inversion effects in the expert classification of mammograms and faces. *Cognitive Research: Principles and Implications*, *3*(1), Article 31. https://doi.org/10.1186/s41235-018-0123-6

Chua, K. W., Richler, J. J., & Gauthier, I. (2014). Becoming a Lunari or Taiyo expert: Learned attention to parts drives holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 1174–1182. https://doi.org/10.1037/a0035895

Chua, K. W., Richler, J. J., & Gauthier, I. (2015). Holistic processing from learned attention to parts. *Journal of Experimental Psychology: General*, *144*(4), 723–729. https://doi.org/10.1037/xge0000063

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, *26*(6), 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 94–107. https://doi.org/10.1037/0096-1523.35.1.94

Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J. R., Massin, P., Erginay, A., Charton, B., & Klein, J. C. (2014). Feedback on a publicly distributed image database: The Messidor database. *Image Analysis & Stereology*, *33*(3), 231–234. https://doi.org/10.5566/ias.1155

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*(2), 107–117. https://doi.org/10.1037/0096-3445.115.2.107

Dror, I. E., Stevenage, S. V., & Ashworth, A. R. S. (2008). Helping the cognitive system learn: Exaggerating distinctiveness and uniqueness. *Applied Cognitive Psychology*, *22*(4), 573–584. https://doi.org/10.1002/acp.1383

Ericsson, K. A. (2017). Expertise and individual differences: The search for the structure and acquisition of experts' superior performance. *WIREs Cognitive Science*, *8*(1–2), Article e1382. https://doi.org/10.1002/wcs.1382

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*(1), 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology*, *15*(2), 154–160. https://doi.org/10.1016/j.conb.2005.03.010

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44. https://doi.org/10.1038/287043a0

Fiumara, G., Flanagan, P., Schwarz, M., Tabassi, E., & Boehnen, C. (2018). *NIST special database 301*. National Institute of Standards and Technology. https://doi.org/10.18434/T4/1502474

Gallego, A. J., Pertusa, A., & Gil, P. (2018). Automatic ship classification from optical aerial images with convolutional neural networks. *Remote Sensing*, *10*(4), Article 511. https://doi.org/10.3390/rs10040511

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: A framework for studying expert object recognition processes. *Vision Research*, *38*(15–16), 2401–2428. https://doi.org/10.1016/S0042-6989(97)00442-2

Goode, M. K., Geraci, L., & Roediger, H. L., III (2008). Superiority of variable to repeated practice in transfer on anagram solution. *Psychonomic Bulletin & Review*, *15*(3), 662–666. https://doi.org/10.3758/PBR.15.3.662

Growns, B., & Martire, K. A. (2020). Human factors in forensic science: The cognitive mechanisms that underlie forensic feature-comparison expertise. *Forensic Science International. Synergy*, *2*, 148–153. https://doi.org/10.1016/j.fsisyn.2020.05.001

Heitman, R. J., Pugh, S. F., Kovaleski, J. E., Norell, P. M., & Vicory, J. R. (2005). Effects of specific versus variable practice on the retention and transfer of a continuous motor skill. *Perceptual and Motor Skills*, *100*(3 Part 2), 1107–1113. https://doi.org/10.2466/pms.100.3c.1107-1113

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54*(4), 1105–1124. https://doi.org/10.1080/713756001

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84. https://doi.org/10.1016/j.plrev.2008.12.001

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*, *46*(2), 395–401. https://doi.org/10.1177/003151257804600201

Kortylewski, A., Albrecht, T., & Vetter, T. (2015). Unsupervised footwear impression analysis and retrieval from crime scene data. In C. Jawahar & S. Shan (Eds.), *Computer vision—ACCV 2014 workshops. ACCV 2014. Lecture notes in computer science* (Vol. 9001, pp. 644–658). Springer. https://doi.org/10.1007/978-3-319-16628-5_46

Landin, D. K., Hebert, E. P., & Fairweather, M. (1993). The effects of variable practice on the performance of a basketball skill. *Research Quarterly for Exercise and Sport*, *64*(2), 232–237. https://doi.org/10.1080/02701367.1993.10608803

Leavey, P., Sengupta, A., Rakheja, D., Daescu, O., Arunachalam, H. B., & Mishra, R. (2019). *Osteosarcoma data from UT Southwestern/UT Dallas for viable and necrotic tumor assessment* [Data set]. Cancer Imaging Archive. https://doi.org/10.7937/tcia.2019.bvhjhdas

Liesefeld, H. R., Fu, X., & Zimmer, H. D. (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed-accuracy trade-offs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1140–1151. https://doi.org/10.1037/xlm0000081

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Nakashima, R., Kobayashi, K., Maeda, E., Yoshikawa, T., & Yokosawa, K. (2013). Visual search of experts in medical image reading: The effect of training, target prevalence, and expert knowledge. *Frontiers in Psychology*, *4*, Article 166. https://doi.org/10.3389/fpsyg.2013.00166

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Nodine, C. F., & Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology*, *5*(9), 603–612. https://doi.org/10.1016/S1076-6332(98)80295-X

Roads, B., Mozer, M. C., & Busey, T. A. (2016). Using highlighting to train attentional expertise. *PLoS ONE*, *11*(2), Article e0149368. https://doi.org/10.1371/journal.pone.0149368

Robson, S. G., Searston, R. A., Edmond, G., McCarthy, D. J., & Tangen, J. M. (2020). An expert–novice comparison of feature choice. *Applied Cognitive Psychology*, *34*(5), 984–995. https://doi.org/10.1002/acp.3676

Robson, S. G., Tangen, J. M., & Searston, R. A. (2021). The effect of expertise, target usefulness and image structure on visual search. *Cognitive Research: Principles and Implications*, *6*(1), Article 16. https://doi.org/10.1186/s41235-021-00282-5

Roller, C. A., Cohen, H. S., Kimball, K. T., & Bloomberg, J. J. (2001). Variable practice with lenses improves visuo-motor plasticity. *Cognitive Brain Research*, *12*(2), 341–352. https://doi.org/10.1016/S0926-6410(01)00077-5

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770

Schuster, D., Rivera, J., Sellers, B. C., Fiore, S. M., & Jentsch, F. (2013). Perceptual training for visual search. *Ergonomics*, *56*(7), 1101–1115. https://doi.org/10.1080/00140139.2013.790481

Scott, L. S., Tanaka, J. W., Sheinberg, D. L., & Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, *1210*, 204–215. https://doi.org/10.1016/j.brainres.2008.02.054

Searston, R. A., & Tangen, J. M. (2017a). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE*, *12*(6), Article e0178403. https://doi.org/10.1371/journal.pone.0178403

Searston, R. A., & Tangen, J. M. (2017b). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review*, *24*(4), 1324–1329. https://doi.org/10.3758/s13423-016-1211-6

Searston, R. A., & Tangen, J. M. (2017c). Training perceptual experts: Feedback, labels, and contrasts. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(1), 32–39. https://doi.org/10.1037/cep0000124

Searston, R. A., Tangen, J. M., & Thompson, M. B. (2021, April). *A multiple-task reduction approach to measuring perceptual expertise in fingerprint analysis* [Paper presentation]. The 47th Experimental Psychology Conference, Brisbane, Australia.

Shea, C. H., & Kohl, R. M. (1990). Specificity and variability of practice. *Research Quarterly for Exercise and Sport*, *61*(2), 169–177. https://doi.org/10.1080/02701367.1990.10608671

Shen, J., Mack, M. L., & Palmeri, T. J. (2014). Studying real-world perceptual expertise. *Frontiers in Psychology*, *5*, Article 857. https://doi.org/10.3389/fpsyg.2014.00857

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176–199. https://doi.org/10.1177/1745691615569000

SWGFAST. (2012). *Standard for the documentation of analysis, comparison, evaluation, and verification Version 2.0*. http://clpex.com/swgfast/Documents.html

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, *16*(2), 145–151. https://doi.org/10.1111/j.0956-7976.2005.00795.x

Tangen, J. M., Thompson, M. B., & McCarthy, D. J. (2011). Identifying fingerprint expertise. *Psychological Science*, *22*(8), 995–997. https://doi.org/10.1177/0956797611414729

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE*, *9*(12), Article e114759. https://doi.org/10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior*, *38*(1), 84–93. https://doi.org/10.1037/lhb0000051

Thompson, M. B., Tangen, J. M., & Searston, R. A. (2014). Understanding expertise and non-analytic cognition in fingerprint discriminations made by humans. *Frontiers in Psychology*, *5*, Article 737. https://doi.org/10.3389/fpsyg.2014.00737

Towler, A., Keshwa, M., Ton, B., Kemp, R. I., & White, D. (2021). Diagnostic feature training improves face matching accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(8), 1288–1298. https://doi.org/10.1037/xlm0000972

Vakil, E., & Heled, E. (2016). The effect of constant versus varied training on transfer in a cognitive skill learning task: The case of the Tower of Hanoi puzzle. *Learning and Individual Differences*, *47*, 207–214. https://doi.org/10.1016/j.lindif.2016.02.009

Vanderkolk, J. R. (2011). Examination process. In A. McRoberts (Ed.), *The fingerprint sourcebook* (pp. 9-3–9-26). U.S. Dept. of Justice, Office of Justice Programs, National Institute of Justice.

Vogelsang, M. D., Palmeri, T. J., & Busey, T. A. (2017). Holistic processing of fingerprints by expert forensic examiners. *Cognitive Research: Principles and Implications*, *2*(1), Article 15. https://doi.org/10.1186/s41235-017-0051-x

Willey, C. R., & Liu, Z. (2018). Long-term motor learning: Effects of varied and specific practice. *Vision Research*, *152*, 10–16. https://doi.org/10.1016/j.visres.2017.03.012

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert—but which type? *Psychological Science*, *20*(9), 1108–1117. https://doi.org/10.1111/j.1467-9280.2009.02430.x

Received October 17, 2021
Revision received June 20, 2022
Accepted July 24, 2022 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.