Journal of Applied Research in Memory and Cognition 6 (2017) 442–451

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

The Emergence of Perceptual Expertise with Fingerprints Over Time

Rachel A. Searston*
The University of Melbourne, Australia

Jason M. Tangen
The University of Queensland, Australia

Can early individual differences in performance predict later expertise in the applied domain of fingerprint identification? We tracked 24 new trainees over the course of a year as they accumulated experience working in a fingerprint unit. We tested their performance every three months on four measures of fingerprint expertise. Trainees significantly improved on all four measures, with the majority of learning occurring within the first three months. When we indexed trainees' performance, by averaging across their percent correct scores on all four measures of expertise, we found early indexed performance was significantly and positively related to their indexed performance three, six, nine, and 12 months later. These findings provide a rich example of how perceptual expertise can emerge within an applied domain, and evidence that early individual differences on a composite measure of performance can be diagnostic of later expertise.

General Audience Summary

How does expertise develop in radiology, face recognition or fingerprint identification? Surprisingly few studies have examined the development of expertise over a long period of time. We also know little about whether some people are more cut out for these applied domains. We addressed this gap in the context of fingerprint identification, by examining the performance of trainee examiners over their first 12 months of working in a fingerprint unit. We tested trainee examiners on four established measures of fingerprint expertise every three months in their workplace, and indexed their performance on each occasion by averaging across their percent correct scores. We found that trainees' accuracy on the fingerprint index (and on each measure separately) improved considerably with just three months experience, but learning plateaued after this time. Trainees' early scores on the fingerprint index were also a reliable predictor of their indexed performance three, six, nine, and 12 months later—meaning that the top performers tended to remain at the top. These findings have implications for theories of perceptual expertise because they provide compelling evidence that both experience and prior individual differences can be diagnostic of performance in an applied perceptual domain. Within the context of fingerprint identification, these findings demonstrate that training and experience in the domain—a benchmark often used to make decisions about the admissibility of expert evidence in legal proceedings—contributes to the development of fingerprint expertise. The development of evidence-based training methods and selection tools could be useful avenues for more efficiently cultivating expert examiners.

Keywords: Perceptual expertise, Individual differences, Forensic science, Expertise acquisition, Work-based learning

Author Note

Jason M. Tangen, School of Psychology, The University of Queensland, Australia.

This research was supported by an Australian Research Council Linkage Grant (LP120100063) to Jason M. Tangen and a University of Melbourne McKenzie Postdoctoral Research Fellowship to Rachel A. Searston.

* Correspondence concerning this article should be addressed to Rachel A. Searston, Melbourne Centre for the Study of Higher Education, Melbourne Graduate School of Education, The University of Melbourne, Parkville, VIC 3010, Australia. Contact: rachel.searston@gmail.com

People vary in their ability to recognise perceptual categories. Developmental prosopagnosics perform significantly below average at recognising faces (Kress & Daum, 2003) and others—termed *super recognisers*—perform significantly above average across face memory and face matching tasks (Bobak, Hancock, & Bate, 2015; Russell, Duchaine, & Nakayama, 2009). Other examples of expert-novice differences in perceptual domains abound (Tarr & Cheng, 2003). The prevailing view in cognitive psychology is that variation in performance is largely a result of variation in the amount of deliberate practice an individual engages in (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Ericsson, 2007, 2014; Ericsson, Krampe, & Tesch-Römer, 1993). Indeed, in perceptual domains experience with identifying and discriminating bird species (Tanaka, Curran, & Sheinberg, 2005), other-race faces (Bukach, Cottle, Ubiwa, & Miller, 2012), shapes (Garrigan & Kellman, 2008), and fictitious beasts (Gauthier & Tarr, 1997; Wong, Palmeri, & Gauthier, 2009), has been shown to facilitate learning and the development of expertise. While deliberate practice is undoubtedly important, commentators have called for research examining whether effects of expertise observed across a wide range of domains could also be underpinned by prior individual differences (e.g., Macnamara, Hambrick, & Oswald, 2014). We test this hypothesis in the perceptual domain of fingerprint identification, by examining whether prior individual differences in performance on a composite measure of fingerprint expertise can predict performance on that same measure over a 12-month period of working in a fingerprint unit.

Dissecting Existing Expertise

A common approach to assessing individual differences in perceptual expertise has been to retrospectively examine the relationship between domain-specific measures of performance and measures of more general abilities, such as IQ and visual memory. Other studies have made use of retrospective twin designs to determine whether there is a genetic component to expertise. In face recognition, there is evidence of a higher correlation between identical twins compared to fraternal twins in their ability to recognise faces (Wilmer et al., 2010), but no correlation between face recognition ability and general intelligence (IQ), or general visual memory (Davis et al., 2011). Further afield, others have reported a genetic component to reading skill using a retrospective twin design (Plomin, Shakeshaft, McMillan, & Trzaskowski, 2014), and a significant relationship between several measures of general cognitive ability (e.g., IQ and visual memory) and chess skill among children (Bilalić, McLeod, & Gobet, 2007; Horgan & Morgan, 1990), and adults (Grabner, Stern, & Neubauer, 2007; but for conflicting results, see Waters, Gobet, & Leyden, 2002). While these findings offer insights about a possible source of variation among individuals, it is impossible to tease apart the relative contribution of experience using retrospective methods. With twins, for instance, fraternal pairs may vary more than identical pairs in their sets of experiences, resulting in more varied performance.

Predicting Future Expertise

A second approach to assessing individual differences in expertise is to predict future achievement based on current performance. One particular domain that has an established literature on predicting future achievement is medicine. Admission to postgraduate medical science programmes is highly competitive, the candidates are all highly qualified, and attrition rates are typically very low (Eva, Rosenfeld, Reiter, & Norman, 2004; Salvatori, 2001). The traditional personal interview approach to selecting candidates is also prone to context specificity effects (Eva et al., 2004). To overcome some of these issues, medical education researchers developed the *multiple mini-interview*, a selection tool that involves averaging across scores from multiple samples of short, structured interviews with candidates (Eva et al., 2004). Increasing the number of interviews (and interviewers) dilutes the chances of candidates being selected on the basis of compatibility with a particular interview panel, or a once off favourable performance on the day. We borrow aspects of this multiple samples method in our current study. Predicting future behaviour, however, also has limits. There is no way to assess the future performance of selected candidates relative to the future performance of rejected candidates *had* they been selected (Schmidt & Hunter, 1998). Because the range of people to test down the track is restricted, it can be difficult to know whether a selection tool actually discriminates top future performers from the rest.

Tracking the Emergence of Expertise Over Time

A third approach to testing whether individual differences underlie perceptual expertise, is to collect longitudinal data. However, there are surprisingly few longitudinal studies mapping the development of expertise over time in applied domains. Prior work integrating dual-process and individual differences theories suggests that early individual differences among naïve performers remain stable over a period of learning (in a single session) when the task demands vary (e.g., randomly intermixing targets and distractors on a verbal category search task), but they diminish when the task demands remain consistent (e.g., colour naming, symbol sorting; see Ackerman, 1987). Tasks with *inconsistent* components are thought to imply the use of more controlled or effortful processes, whereas tasks with *consistent* components are thought to become more automated with experience. From this perspective, individual differences in general ability may be equated with differences in cognitive capacity or amount of attentional resources, and the transition from controlled to more automatic processing with expertise is synonymous with becoming less sensitive to general resource limitations (Ackerman, 1987). While this work is based on relatively artificial cognitive tasks, it offers a theoretical framework for assessing individual differences in perceptual expertise. Early individual differences in learners' ability to classify and discriminate objects or categories might remain stable for tasks that are inconsistent, novel, or that allow controlled, effortful, analytic processing, but not for tasks that come to rely on fast, intuitive, non-analytic processing with experience.

The Present Study: Tracking the Emergence of Expertise with Fingerprints Over Time

In this study, we probe whether early individual differences can be diagnostic of later perceptual expertise in the context of fingerprint identification. Specifically, we test whether variation in performance among a group of new fingerprint trainees predicts variation in their performance as they accumulate experience with fingerprints in a natural setting: working a fingerprint unit (see SWGFAST, 2012b, for an example field training standard). We measure their performance every three months, for 12 months, on a suite of perceptual tasks previously demonstrated to distinguish between expert and novice fingerprint examiners. These tasks included a self-paced visual search (Searston & Tangen, 2017b) and fingerprint matching task (Tangen, Thompson, & McCarthy, 2011), a fingerprint matching task where the images are presented very briefly (Searston & Tangen, 2017b; Thompson & Tangen, 2014), and a fingerprint matching task where we shift the level of specificity from the finger to the person (Searston & Tangen, 2017a).

Fingerprint examiners spend their days matching *unfamiliar* images. Their task is to determine whether two prints they have never seen before, belong to the same unfamiliar finger (e.g., different instances of Smith's right thumb), or two different unfamiliar fingers (e.g., instances of Smith's right thumb and Jones's right thumb). The novelty of each case could be considered a form of inconsistency (Ackerman, 1987), and performance under these conditions may therefore require effortful processing that is dependent on more general cognitive abilities. There is evidence to suggest that fingerprint examiners rely somewhat on controlled, effortful or analytic processes when matching prints. The difference in fingerprint matching performance between experts and novices is greater when given 60 s of viewing time versus one-second (Thompson & Tangen, 2014), and formal practice guidelines encourage a slow analytic process of marking up and comparing particular features in each case before arriving at a conclusion (SWGFAST, 2012a). From this perspective, we would expect individual differences in performance on measures of fingerprint expertise to remain stable over time, irrespective of experience. In practice, such a result would suggest that composite measures, and other multiple-sample methods for assessing performance, may hold promise as tools for selecting new trainees in fingerprints.

There is also evidence that fingerprint examiners make use of non-analytic or automatic processes with expertise. Experts are more accurate than novices with matching prints that are spaced briefly in time or presented briefly on the screen (Thompson & Tangen, 2014), and they display physiological characteristics of configural processing (Busey & Vanderkolk, 2005). Fingerprint experts are also able to discriminate prints belonging to different fingers of the same person (e.g., Smith's right thumb and little finger; Searston & Tangen, 2017a), and fingerprint matching decisions appear to be influenced by similarity to prior cases (Searston, Tangen, & Eva, 2016), suggesting a reliance on memory and a feeling of familiarity when matching unfamiliar images. Based on this evidence, we would expect prior individual differences in performance on measures of fingerprint expertise to diminish as examiners gain experience in the domain. Such a result would be consistent with exemplar models that emphasise a greater on reliance on automated memory retrieval with increasing experience (Brooks, 1978; Logan, 1988; Medin & Schaffer, 1978). In practice, this result would suggest that the emphasis currently placed on "training and experience" as an indicator of expertise in forensic science may hold true in the case fingerprints (Busey & Parada, 2010).

Method

Participants

Participants were 24 fingerprint trainees from four police organisations in Australia (Queensland, The Australian Federal, New South Wales, and Victoria Police). Twelve trainees—*Trainee Group A*—had less than two weeks of formal experience with discriminating prints on the first day of testing. The remaining 12 trainees—*Trainee Group B*—had between one and three months experience with matching fingerprints on the first day of testing (*Mean Experience* = 1.7 months). This sequential strategy (Schaie, 1965) of examining changes within groups over time and between groups with different levels of experience on the first day of testing, enabled us to better isolate whether any change in performance was a genuine effect of trainees' experience working in a fingerprint unit, or a result of being tested. Trainees were spread over six small cohorts or intakes, and typically started in their jobs within days of each other in each cohort. They were recruited during three-monthly visits to each of the police organisations over a three-year period. Given the nature of their work and relatively small populations, similar sample sizes are typical of studies involving fingerprint examiners (e.g., Searston & Tangen, 2017a; Tangen et al., 2011; Vogelsang, Palmeri, & Busey, 2017) and experts in other applied domains (e.g., Brooks, Norman, & Allen, 1991; Towler, White, & Kemp, 2017). Nonetheless, previous studies comparing expert fingerprint examiners to novices on the tasks used in the current study have reported moderate to large effect sizes (Searston & Tangen, 2017a, 2017b).

Five of the 24 trainees were not able to complete all five sessions: two (one from Trainee Group A, and one from Trainee Group B) were not available at nine or 12 months, one (from Trainee Group A) was not available at nine months, one (from Trainee Group B) was not available at six months, and another (from Trainee Group B) was only available to complete the tasks on the first visit. Some missing data due to leave arrangements was unavoidable, and we have included the data we obtained from these trainees in our analyses where possible.

While we were unable to keep a thorough record of trainees' experiences, the first author completed a week of fingerprint training alongside trainees in one organisation. This training involved a mix of theory (e.g., learning about the biology of friction ridge skin, concepts like distortion, and causal knowledge about why a fingerprint may change from one time to the next) and practice (e.g., classifying and comparing fingerprints). Training to become a fingerprint expert in Australia typically spans a five-year period, during which time trainees are mentored by more experienced examiners on the job, participate in structured lessons, and complete theoretical and practical examinations.

Visual Search

As one measure of fingerprint expertise, trainees viewed 40 arrays of 40 prints one at a time on the computer screen, and were asked to find and select the whorl among the loops or the loop among the whorls as quickly and as accurately as possible (see Figure 1a for an example whorl array with a loop target to the left). Prior work has shown that fingerprint experts are significantly more accurate than novices at locating categorical outliers in these arrays (Searston & Tangen, 2017b). Trainees also completed an inverted face version of the same task—where trainees viewed 40 arrays of 40 inverted faces, selecting the inverted male face among the inverted female faces or vice versa (see to the right of Figure 1a). Our expertise with faces is disrupted when the images are inverted (Young, Hellawell, & Hay, 1987), and we expected trainees to become more accurate on the fingerprint task (but not the inverted face task) as they gained experience with matching prints.

The prints were 200 fingerprint patterns (100 loops and 100 whorls), collected from different fingers of 30 individuals. Each print was cropped to 180 × 180 pixels and we applied a circular mask. Forty arrays were generated for each trainee: 20 with a loop pattern among whorl distractors, and 20 with a whorl pattern among loop distractors. The 39 distractors in each array were randomly sampled from the remaining pool of prints. No target was repeated, the position of the target was random in each array, and the sequence of trials (i.e., loop or whorl target) was random for each trainee each time the task was completed. We used the same method to generate 40 arrays of 40 inverted face images with inverted male or inverted female target images. The faces were a subset of 100 male and 100 female photographs from the Face Recognition Grand Challenge database (Phillips et al., 2005), cropped to 180 × 180 pixels, inverted, and masked so that the hair could not be seen.

Speeded Matching

Prior work has shown also that fingerprint experts are more accurate than novices when matching prints presented for one second (Thompson & Tangen, 2014), and just 400 ms (Searston & Tangen, 2017b). As a second measure of fingerprint expertise, trainees viewed pairs of prints for 400 ms followed by a 50 ms visual mask (i.e., scrambled images of the fingerprints), and were asked to judge whether the prints were left by the same finger or two different fingers (see to the left of Figure 1b for an example task sequence). There were 100 pairs (50 matching and 50 mismatching), and trainees indicated their judgments on a 12 point, forced-choice confidence rating scale (Tangen et al., 2011). We also included an inverted face version of this task, where trainees viewed pairs of photographs for 400 ms followed by a 50 ms visual mask. In this version, they were asked to judge whether the photographs depicted the same individual or two different individuals (see to the right of Figure 1b for an example task sequence). Again, we expected trainees to become more accurate at discriminating fingerprints (but not inverted faces) as they gained experience working with fingerprints.

The prints were the same set used in the visual search task without the circular mask. One-hundred target prints were randomly paired with either a matching print recorded on a separate occasion (creating 50 matching pairs), or a mismatching print from a random other individual in the set (creating 50 mismatching pairs). The images were cropped to 675 × 675 pixels with the prints isolated in the centre. The order of matching and mismatching trials was random for each trainee on each occasion they completed the task. The faces were also the same as those used in the visual search task, only the hair was not masked and the images were cropped to 675 × 900 pixels.

Person Matching

Another facet of fingerprint expertise is an ability to detect stylistic, relational or family resemblance information among a person's prints. Fingerprint experts are more accurate than novices at judging whether a series of prints was left by the same *person*, even when they were left by completely different fingers (e.g., Jones's right thumb, index, middle, ring and little fingerprints are all instances of Jones; Searston & Tangen, 2017a). As a third measure of fingerprint expertise, trainees judged whether lineups of five prints, each from different fingers (i.e., index, middle, ring, little, and thumb), were left by the same *person* or whether one of the prints was left by another individual (see Figure 1c for an example matching and mismatching fingerprint lineup). The lineups remained on the computer screen until trainees indicated their judgement, using a 12-point forced choice confidence rating scale. This scale ranged from 1 (*sure different*) to 12 (*sure same*), where ratings of 1–6 indicated a "no match" decision (i.e., they thought one of the prints belonged to another individual) and ratings 7–12 indicated a "match" decision (i.e., they thought all five impressions belonged to the same individual). The image set used in this task consisted of 10 fully rolled prints from 60 individuals that were cropped to 600 × 600 pixels, and masked to isolate the structure of the prints (Searston & Tangen, 2017a). Sixty lineups of five prints were generated for each trainee on each testing occasion. The lineups were sampled equally from the left and right hand and further partitioned equally as matching and or mismatching trials. The distractors were sampled from a random other individual in the set while still ensuring that all lineups included a print of each of the five digit types. The sequence of matching and mismatching trials was also random for each trainee on each testing occasion.

Fingerprint Matching

The final measure of expertise was a latent fingerprint matching task that has also been shown to discriminate between experts and novices (Tangen et al., 2011). Trainees viewed 36 pairs of prints (18 matching and 18 mismatching) and judged them as belonging to the *same* finger or two *different* fingers—the fingerprint pairs remained on the computer screen until trainees indicated their judgement. The prints were a subset of those used

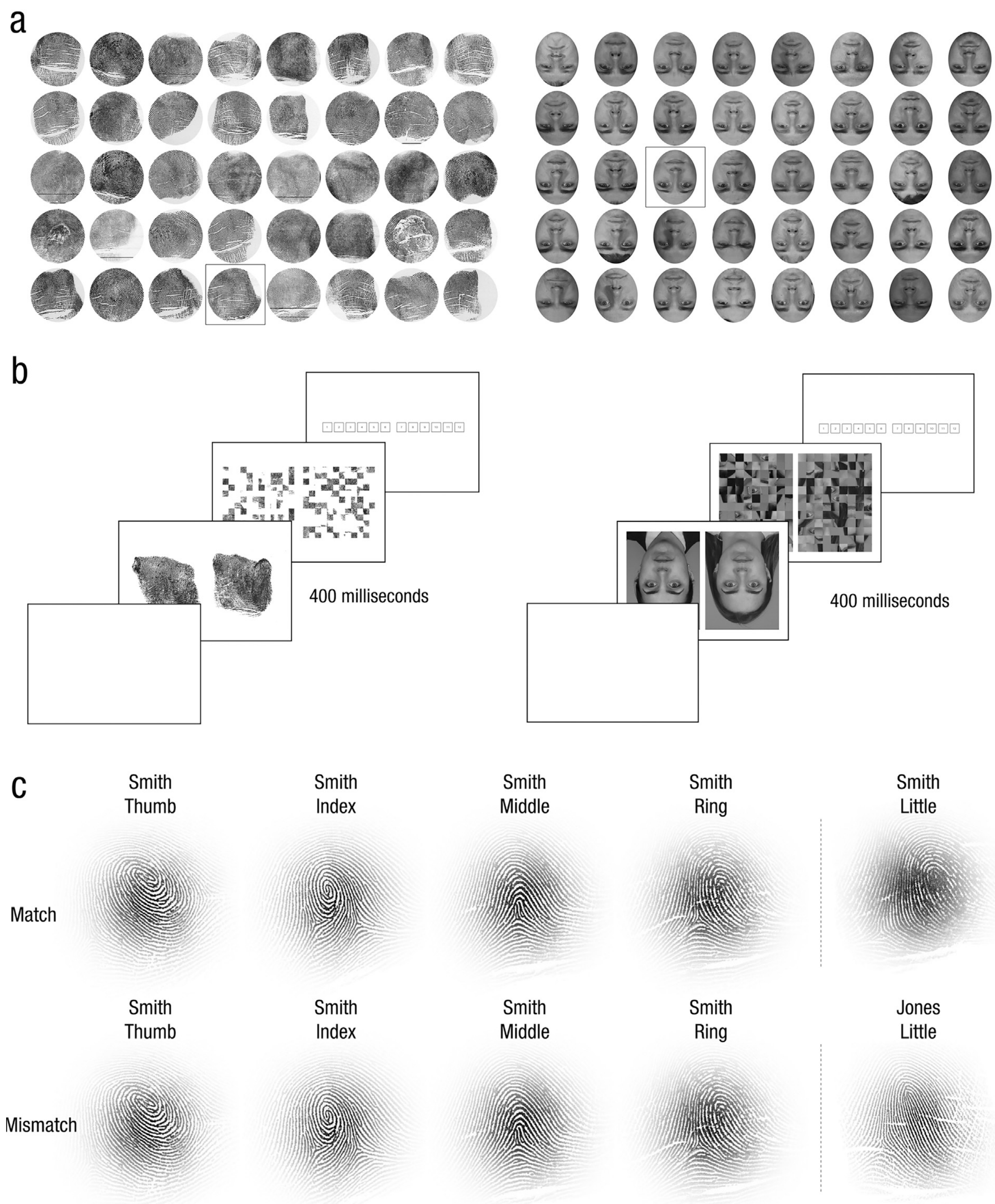

Figure 1. An example fingerprint array with a loop target (a left), and an inverted face array with a female target (a right). (b) depicts the speeded fingerprint matching task sequence (b left), and the speeded inverted face matching task sequence (b right). (c) depicts an example matching and mismatching fingerprint lineup in the person matching task.

in Tangen et al. (2011); we did not include a random non-match condition. The 36 fingerprint pairs were generated by randomly sampling 18 simulated crime-scene prints, pairing them with a corresponding fully rolled print left by the same finger, and pairing the remaining 18 with a corresponding highly similar but mismatching print.

Procedure

All trainees read an information sheet about the study, and were told the aim of the experiment was to track their development as they gained experience (with fingerprints). They also watched an instructional video about each task, with a reminder at beginning of each testing session. The suite of measures was presented in a different randomised order on five separate occasions, spaced three months apart. They were presented on a Macbook Air in a room separate from the trainees' workspaces, and each testing session took approximately two hours to complete with short breaks between the tasks. On some occasions, the trainees completed the tasks over two separate days, depending on their work demands.

Results

Effect of Experience on Trainees' Indexed Performance

Trainees improved on all four measures of fingerprint expertise as they gained experience working in a fingerprint unit. We indexed each trainee's performance by averaging over their percent correct scores on the four fingerprint tasks (excluding the two inverted faces tasks). See Figure 2, for an illustration of each trainee's indexed performance over the 12 months (the light and dark turquoise lines indicate the average percent correct for Trainee Group A and Trainee Group B). On the first day of testing, Trainee Group A averaged 66.4% (SD = 11.5%) correct on the index, improving to 78.6% (SD = 6.4%) with just three months experience with matching prints. This improvement remained steady with six (78.7% correct; SD = 6.9%), nine (79.5%; SD = 7.7%), and 12 months experience (80.3%; SD = 6.3%). Trainee Group B showed a similar pattern of results, averaging 79.1% (SD = 5.1%) correct on the index with one to three months experience, increasing to 80.7% (SD = 4.9%), 83.5% (SD = 4.8%), and then 84.2% (SD = 2.6%) three, six, and nine months later.

We subjected Trainee Group A's index scores to a within-subjects one-way analysis of variance (Experience: none, three months, six months, nine months, and 12 months), finding a significant overall effect of Experience on trainees' average percent correct scores across the four fingerprint tasks, $F(1, 9) = 26.85$, $p < .001$, $\eta^2_G = .75$. Follow up paired comparisons revealed that these trainees improved significantly with three months experience working in a fingerprint unit, $t(11) = 5.763$, $p < .001$, Cohen's $d = 1.31$, and this improvement plateaued from three to six, $t(11) = .11$, $p = .913$, three to nine, $t(9) = 1.02$, $p = .333$, and 3–12 months, $t(9) = 1.91$, $p = .089$. We then ran a second one-way analysis of variance between-subjects, substituting Trainee Group A's index scores at 3, 6, 9, and 12 months with Trainee Group B's scores at those equivalent times. Again, we found a significant effect of Experience, $F(1, 52) = 34.12$, $p < .001$, $\eta^2_G = .40$. With one to three months experience, Trainee Group B outperformed Trainee Group A, when they were at the beginning of their training, $t(46) = 3.49$, $p < .003$, Cohen's $d = 1.43$. There was no discernible improvement in Trainee Group B's performance on the index from three to six months, $t(44) = .74$, $p = .462$, or from three to nine months, $t(44) = 2.09$, $p = .050$, though they did show a significant improvement from three to 12 months, $t(44) = 3.06$, $p = .007$, Cohen's $d = 1.26$. See Supplemental Material for detailed analyses of the trainees' performance on each of the four fingerprint measures, and two face measures separately. Within-subjects analyses of trainees' performance on the two face tasks revealed no significant improvements.

Individual Differences on the Fingerprint Index

Collapsing across all four fingerprint tasks, early individual differences remained diagnostic of performance over the 12 months. The percent correct scores for Trainee Group A on the fingerprint index at the beginning of their training were significantly correlated with their performance on the index three, $r(12) = .82$, $p = .001$, six, $r(12) = .77$, $p = .003$, nine, $r(10) = .79$, $p = .007$, and 12 months later, $r(11) = .72$, $p = .012$ (see Figure 3 for scatterplots of these results). Of primary interest was whether individual differences on the fingerprint index without any training or experience in the domain were predictive of individual differences as trainees' gained experience. We were unable to perform the same individual differences analysis with Trainee Group B, as they already had one to three months experience working in a fingerprint unit when they first completed the tasks. Their scores on the fingerprint index with one to three months experience were positively and significantly correlated with their performance at nine months, $r(10) = .79$, $p = .007$, but not six, $r(10) = .62$, $p = .056$ or 12 months, $r(10) = .50$, $p = .141$.

Individual Differences on the Visual Search and Fingerprint Matching Tasks

Early individual differences in performance also remained stable over the 12 months when we removed the speeded and person matching tasks from the index. We calculated the average percent correct for each trainee collapsed across the fingerprint visual search and fingerprint matching tasks. The average scores for Trainee Group A across these two measures when they had no experience working in a fingerprint unit, were significantly correlated with their performance on the same two measures three, $r(12) = .64$, $p = .025$, six, $r(12) = .62$, $p = .033$, nine, $r(10) = .73$, $p = .016$, and 12 months later, $r(11) = .75$, $p = .008$. When we ran these analyses using data from the fingerprint matching task only, however, we found no significant correlation between trainees' fingerprint matching accuracy at the beginning of their training and their fingerprint matching accuracy three, $r(12) = .51$, $p = .093$, or six months later, $r(12) = .40$, $p = .193$. Their earliest performance on the fingerprint matching task was significantly related to their fingerprint matching accuracy after nine months of fingerprint training, $r(10) = .68$, $p = .032$, but not after 12 months, $r(11) = .60$, $p = .051$. Early performance on the fingerprint visual search task by itself was not significantly correlated with trainees' performance on the same task after three, $r(12) = .49$, $p = .103$, six, $r(12) = .50$, $p = .095$, or nine months experience with matching prints, $r(10) = .63$, $p = .053$, but it was after 12 months, $r(11) = .67$, $p = .024$.

Individual Differences on the Speeded and Person Matching Tasks

For the speeded and person matching tasks, early individual differences became less predictive of performance as trainees gained experience. Collapsing over the two measures, the performance of Trainee Group A at the beginning of their training significantly predicted their performance on the same combined measure three months later, $r(12) = .61$, $p = .035$, but not six,

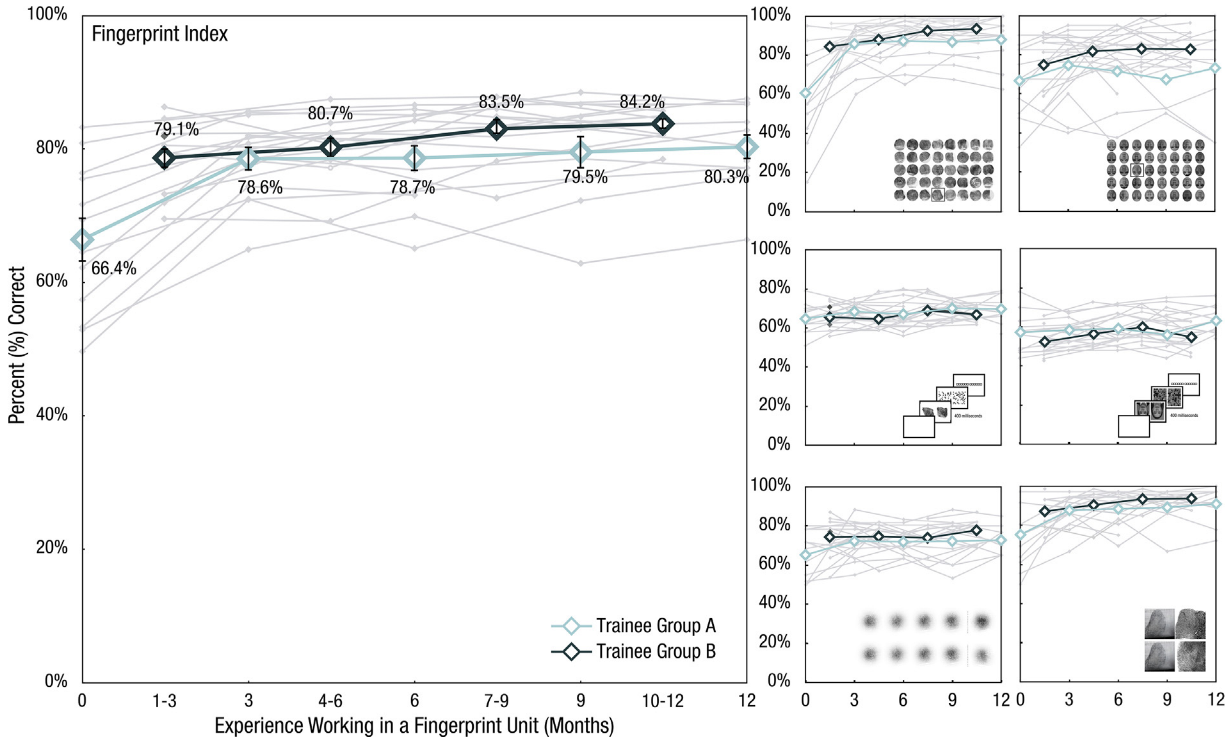

Figure 2. To the left is the mean percent correct for each individual trainee (light grey), for Trainee Group A (dark turquoise), and for Trainee Group B (light turquoise) on the fingerprint index (average percent correct scores collapsed across the four fingerprint tasks) over their first 12 months of working in a fingerprint unit. Error bars represent the standard error of the mean, and Trainee Group B has been plotted in-between Trainee Group A to illustrate the one to three month lag before they first completing the suite of tasks. To the right of the index is the mean proportion correct for each individual, Trainee Group A, and Trainee Group B on the fingerprint search task (top left), inverted face search task (top right), speeded fingerprint matching task (middle left), speeded inverted face matching task (middle right), person matching task (bottom left), and fingerprint matching task (bottom right).

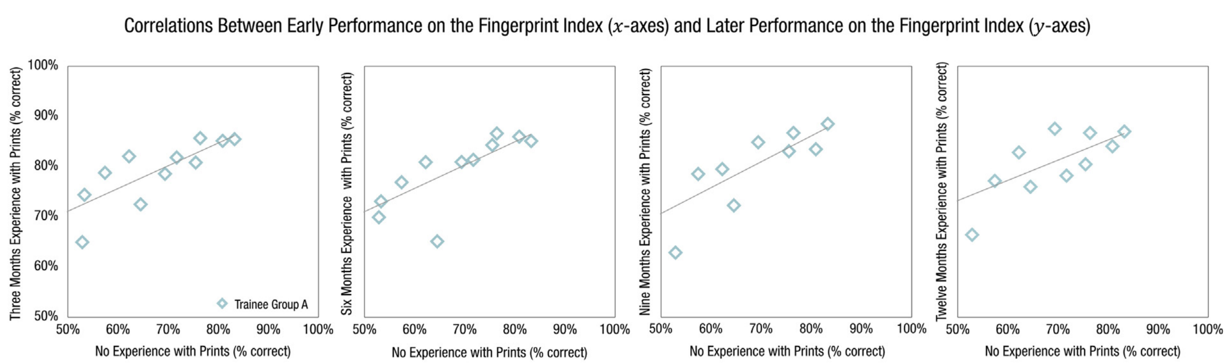

Figure 3. Mean percent correct scores for each trainee (represented individually by the light turquoise diamonds) in Trainee Group A on the fingerprint index (average of percent correct scores collapsed across the four fingerprint tasks) with no experience matching prints along the *x*-axes, and their percent correct scores on the fingerprint index with three, six, nine, and 12 months experience along the *y*-axes.

$r(12) = .44$, $p = .152$, nine, $r(10) = .49$, $p = .151$, or 12 months later, $r(11) = .15$, $p = .660$. A similar pattern emerged when we examined individual differences in performance for each of these tasks separately. Trainees' accuracy with discriminating briefly presented prints when they had no experience in fingerprints whatsoever, was significantly related to their performance on the same task after they had accumulated three months experience with matching fingerprints, $r(12) = .59$, $p = .045$, but it was not related with their performance after six, $r(12) = .37$, $p = .232$, nine, $r(10) = .60$, $p = .068$, or 12 months, $r(11) = .39$, $p = .238$. The accuracy with discriminating people at the beginning of their training was not significantly related to their ability to discriminate people after three, $r(12) = .33$, $p = .355$, six, $r(12) = .37$, $p = .298$, nine, $r(10) = .17$, $p = .643$, or 12 months of working in fingerprints, $r(11) = -.04$, $p = .912$.

Discussion

Few studies have examined the development of expertise over an extended period of time. The purpose of this study was to glean a richer understanding of how people develop expertise over time in an applied perceptual domain: fingerprint identification. We also set out to gauge whether early individual differences *can* be diagnostic of performance as people gain experience in their domain of expertise. We tested 24 fingerprint trainees on four specific measures of fingerprint expertise—a visual search, speeded matching, person matching, and fingerprint matching task—every three months over the course of a year as they accumulated experience with fingerprints in a natural setting. We gauged their performance by collapsing across their percent correct scores on all four tasks.

One clear result was that domain-specific experience facilitated the development of perceptual expertise with fingerprints. Over their first 12 months working in a fingerprint unit, trainees displayed a significant improvement in their ability to: identify categorical outliers (e.g., locating a loop fingerprint pattern in an array of whorls), discriminate briefly presented prints, discriminate family resemblances among a person's prints, and discriminate prints when no time constraints were imposed on them. The vast majority of learning also occurred within three months of working in the domain, and trainees with greater experience working in a fingerprint unit when first completing our suite of tasks (Trainee Group B) scored significantly higher on the fingerprint index than trainees with no experience (Trainee Group A). The significant change in performance *within* trainees over three months, and the significant difference *between* two groups of trainees, who were either new or had been working in a fingerprint unit for one to three months the first time we tested them, provides compelling converging evidence that experience in the domain enhances the development of fingerprint expertise. These results are in line with prior work suggesting that expertise is facilitated by deliberate practice in the domain (Charness et al., 2005; Ericsson, 2007, 2014; Ericsson et al., 1993), and with exemplar models of categorisation and automaticity that posit a greater reliance on a memory for similar instances with experience (Brooks, 1978; Logan, 1988; Medin & Schaffer, 1978). Practically, this finding suggests that experience working in a fingerprint unit contributes to the development of expertise with fingerprints, lending empirical support to the "training and experience" heuristic used in legal proceedings to help judge the admissibility of expert fingerprint evidence (Busey & Parada, 2010).

Perhaps most interesting is the result that trainees' indexed performance on the suite of fingerprint tasks at the beginning of training was significantly related to their overall performance three, six, nine, and 12 months later. That is, the individual differences trainees brought to bear when they had limited experience with fingerprints were diagnostic of their relative performance throughout their first year of training; the top performers tended to remain at the top. This finding suggests that some variation in perceptual expertise (with fingerprints) may be dependent on preexisting differences between individuals (Macnamara et al., 2014), and that it may be possible to design selection tools that are diagnostic of fingerprint expertise down the track. However, the shrinking variance between individuals over the 12 months hints at the possibility that the predictive power of our composite measure may wash out over a longer period of time. If this is true, then initial individual differences on the fingerprint index may be diluted with enough specific experience in the domain—in this case experience in a fingerprint unit. Practically, managers might then place greater emphasis on training rather than selection if the lowest performers are likely to improve over time. In any case, existing theoretical accounts of expertise (e.g., Ericsson et al., 1993), categorisation (e.g., Brooks, 1978), and automaticity (e.g., Logan, 1988) typically do not feature individual differences, and further research testing the generality of our findings across domains would help to refine our basic understanding of expertise.

We were constrained by a small population of trainee fingerprint examiners, and our findings do not allow us to distinguish whether trainees were adopting a controlled or an automatic process. However, the data do provide some clues as to the sorts of conditions where performance may be more likely to depend on preexisting individual differences. When we split the fingerprint index in two, for example, early performance predicted later performance when we combined the self-paced fingerprint matching and visual search tasks but not when we combined the speeded and person matching tasks. This pattern of results may be reconciled with a dual-process account of individual differences (Ackerman, 1987), which predicts that performance is more likely to be constrained by general cognitive resource limitations in tasks that encourage analytic processing. For example, examiners may be encouraged to adopt a more controlled and analytic process in cases where the prints are particularly novel, counterintuitive, or noisy, and there are few pressures to make a decision quickly. In these cases, a dual-process account predicts that preexisting individual differences would more strongly correlate with examiners' performance.

In the present study, we traded a degree of control to test trainees in their natural setting as they developed perceptual expertise with fingerprints: we tested fingerprint trainees' performance across a range specific tasks on five separate occasions over the course of a year. The multiple-samples longitudinal design we adopted in the current study on fingerprint expertise may serve as a useful model for investigating perceptual expertise and how it develops over time more broadly. Taking multiple samples of people's performance across several tasks and time points provides a measure of expertise that is less susceptible to fluctuations due to independent error factors (e.g., once off performances on the day; Eva et al., 2004). Importantly, it also enables us to test the stability of individual differences over some period of training or practice in a particular domain. As we have shown within our sample of trainees, individual differences in fingerprint expertise were much more stable when we computed the average performance of each individual across tasks, compared to tracking their performance on each task separately. More longitudinal research across domains would compliment existing expert-novice studies on the nature of expertise by providing a richer understanding of how it develops over time. Future longitudinal investigations may manipulate specific aspects of training and experience, for example, which would help to better understand the role of individual differences and how best to create perceptual expertise.

Conflict of Interest Statement

The authors declare no conflict of interest.

Author Contributions

Both authors conceived and designed the study, and analysed and interpreted the data. The first author collected the data and drafted the paper. The second author provided critical revisions.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jarmac.2017.08.006.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*(1), 3–27. http://dx.doi.org/10.1037/0033-2909.102.1.3

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2015). Super-recognisers in action: Evidence from face-matching and face memory tasks. *Applied Cognitive Psychology*, http://dx.doi.org/10.1002/acp.3170 (advance online publication)

Bilalić, M., McLeod, P., & Gobet, F. (2007). Does chess need intelligence? A study with young chess players. *Intelligence*, *35*, 457–470. http://dx.doi.org/10.1016/j.intell.2006.09.005

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch, & B. Lloyd (Eds.), *Cognition and categorisation*. Hillsdale, NJ: Erlbaum.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, *120*(3), 278–287. http://dx.doi.org/10.1037/0096-3445.120.3.278

Bukach, C. M., Cottle, J., Ubiwa, J., & Miller, J. (2012). Individuation experience predicts other-race effects in holistic processing for both Caucasian and Black participants. *Cognition*, *123*(2), 319–1324. http://dx.doi.org/10.1016/j.cognition.2012.02.007

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin & Review*, *17*(2), 155–160. http://dx.doi.org/10.3758/PBR.17.2.155

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

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, *19*, 151–165. http://dx.doi.org/10.1002/acp.1106

Davis, J. M., McKone, E., Dennett, H., O'Connor, K. B., O'Kearney, R., & Palermo, R. (2011). Individual differences in the ability to recognise facial identity are associated with social anxiety. *PLoS ONE*, *6*, e28800. http://dx.doi.org/10.1371/journal.pone.0028800

Ericsson, K. A. (2007). Deliberate practice and the modifiability of the body and mind: Toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology*, *38*, 4–34. http://dx.doi.org/10.1111/j.1467-8721.2007.00533.x

Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, *45*, 81–103. http://dx.doi.org/10.1016/j.intell.2013.12.001

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*, 363–406. http://dx.doi.org/10.1037/0033-295X.100.3.363

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: The multiple mini-interview. *Medical Education*, *38*, 314–326. http://dx.doi.org/10.1046/j.1365-2923.2004.01776.x

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682. http://dx.doi.org/10.1016/S0042-6989(96)00286-6

Garrigan, P. B., & Kellman, P. J. (2008). Perceptual learning depends on perceptual constancy. *Proceedings of the National Academy of Sciences of United States of America*, *105*(6), 2248–2253. http://dx.doi.org/10.1073/pnas.0711878105

Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, *124*, 398–420. http://dx.doi.org/10.1016/j.actpsy.2006.07.008

Horgan, D. D., & Morgan, D. (1990). Chess expertise in children. *Applied Cognitive Psychology*, *4*, 109–128. http://dx.doi.org/10.1002/acp.2350040204

Kress, T., & Daum, I. (2003). Developmental prosopagnosia: A review. *Behavioral Neurology*, *14*(3–4), 109–121. http://dx.doi.org/10.1155/2003/520476

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527. http://dx.doi.org/10.1037/0033-295X.95.4.492

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, *25*(8), 1606–1618. http://dx.doi.org/10.1177/0956797614535810

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. http://dx.doi.org/10.1037/0033-295X.85.3.207

Phillips, J. P., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., et al. (2005). Overview of the face recognition grand challenge. *IEEE Computer Vision and Pattern Recognition*.

Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise. *Intelligence*, *45*, 46–59. http://dx.doi.org/10.1016/j.intell.2013.06.008

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252–257. http://dx.doi.org/10.3758/PBR.16.2.252

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Science Education*, *6*, 159–175. http://dx.doi.org/10.1023/A:1011489618208

Schaie, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, *164*(2), 92–107. http://dx.doi.org/10.1037/h0022371

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. http://dx.doi.org/10.1037/0033-2909.124.2.262

Searston, R. A., & Tangen, J. M. (2017a). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review*, *24*(1), 1324–1329. http://dx.doi.org/10.3758/s13423-016-1211-6

Searston, R. A., & Tangen, J. M. (2017b). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLOS ONE*, http://dx.doi.org/10.1371/journal.pone.0178403

Searston, R. A., Tangen, J. M., & Eva, K. W. (2016). Putting bias into context: The role of familiarity in identification. *Law and Human Behavior*, *40*(1), 50–64. http://dx.doi.org/10.1037/lhb0000154

SWGFAST: Scientific Working Group on Friction Ridge Analysis, Study and Technology. (2012a). *Standard for the documentation of analysis, comparison, evaluation, and verification version 2.0.*. Available at:. http://www.swgfast.org/Documents.html

SWGFAST: Scientific Working Group on Friction Ridge Analysis, Study and Technology. (2012b). *Standards for minimum qualifications and training to competency for friction ridge examiner trainees (latent/tenprint).*. Available at:. http://clpex.com/swgfast/documents/qualifications-competency/121124_Qualifications-Training-Competency_2.0.pdf

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world, perceptual expertise. *Psychological Science*, *16*, 145–151. http://dx.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. http://dx.doi.org/10.1177/0956797611414729

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47–58. http://dx.doi.org/10.1037/xap0000108

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends in Cognitive Sciences*, *7*(1), 23–30. http://dx.doi.org/10.1016/S1364-6613(02)00010-4

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), 1–23. http://dx.doi.org/10.1371/journal.pone.0114759

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), 15. http://dx.doi.org/10.1186/s41235-017-0051-x

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuo-spatial abilities in chess players. *British Journal of Psychology*, *30*, 303–311.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of United States of America*, *107*, 5238–5241. http://dx.doi.org/10.1073/pnas.0913053107

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*, 1108–1117. http://dx.doi.org/10.1111/j.1467-9280.2009.02430.x

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–759. http://dx.doi.org/10.1068/p160747

Received 23 February 2017;
received in revised form 21 August 2017;
accepted 24 August 2017
Available online 17 October 2017