Received: 16 September 2019 | Revised: 12 March 2020 | Accepted: 7 April 2020
DOI: 10.1002/acp.3676

RESEARCH ARTICLE

WILEY

An expert–novice comparison of feature choice

Samuel G. Robson[1] | Rachel A. Searston[2] | Gary Edmond[3] | Duncan J. McCarthy[4] | Jason M. Tangen[1]

[1]School of Psychology, The University of Queensland, Brisbane, Queensland, Australia
[2]School of Psychology, The University of Adelaide, Adelaide, South Australia, Australia
[3]School of Law, University of New South Wales, Sydney, New South Wales, Australia
[4]Forensic Services Branch, Queensland Police Service, Brisbane, Queensland, Australia

Correspondence
Samuel G. Robson, School of Psychology, The University of Queensland, Brisbane, Queensland, Australia.
Email: s.robson@uq.edu.au

Funding information
Australian Research Council, Grant/Award Number: LP170100086

Summary

Perceptual experts have learned to rapidly and accurately perceive the structural regularities that define categories and identities within a domain. They extract important features and their relations more efficiently than novices. We used fingerprint examination to investigate expert–novice differences in feature choice. On each fingerprint within our set, experts and novices selected one feature they thought was most useful for distinguishing a particular print and one feature they thought was least useful. We found that experts and novices often differed in the features they chose, and experts tended to agree more with each other. However, any such expert–novice difference appeared to depend on the image at hand typically emerging when salient or more conspicuous features of a fingerprint were unclear. We suggest that perceptual training ought to direct attention to useful features with the understanding that what is useful may change depending on the clarity of the stimuli.

KEYWORDS
feature selection, fingerprints, forensic science, perceptual expertise, visual expertise

1 | INTRODUCTION

Our ability to discriminate between people and between objects in our visual environment is crucial to our everyday functioning. We become more attuned to the features that distinguish increasingly specific classes or "domains" of stimuli with experience and extended practice (Tanaka, Curran, & Sheinberg, 2005). This kind of perceptual learning is the reason experts can make quick and accurate decisions in their domain (Kellman & Garrigan, 2009). Expert performance in chess (Chase & Simon, 1973), directing air traffic (Seamster, Redding, Cannon, Ryder, & Purcell, 1993), examining X-rays (e.g. Sowden, Davies, & Roling, 2000) or matching fingerprints (Tangen, Thompson, & McCarthy, 2011) relies on finely tuned perceptual discrimination of important features. To perform well in domains like these, experts must identify features that co-occur regularly with different response options to determine the most appropriate decision or course of action (Roads, Mozer, & Busey, 2016). That is, they must effectively select features that enable them to distinguish one category or identity from another. The present study aims to examine whether experts differ from novices in the features they consider to be most and least useful.

The domain of forensic fingerprint examination is ideal for investigating how experience shapes our perception naturally, because fingerprints are a novel class of stimuli to most people, except fingerprint examiners. Contrary to popular television crime shows, fingerprint examination – the process of matching or comparing a crime scene fingerprint ("latent") to a suspect fingerprint and deciding if they belong to the same finger or not – is carried out by a human examiner, not a computer algorithm. With their years of experience, these examiners have become highly familiar with the features of fingerprints that are most regular across impressions of the same finger and most informative for the task of distinguishing people (Searston & Tangen, 2017b; Thompson, Tangen, & Searston, 2014). Like other domains, expertise in fingerprint comparison likely depends on perceptual learning processes that lead to the discovery of important features and their relations. However, relatively little is known about how examiners choose features when examining fingerprints. From an applied perspective, insights about expert–novice differences in fingerprint feature selection could help to identify perceptual processes that can be trained or selected for more efficiently. From a theoretical perspective, a comparison of expert fingerprint examiners and novices can serve as a rich ecological case study of how the perception of the relative usefulness of features may change with extensive experience.

In recent decades, the reliability and validity of many forensic science techniques, including fingerprint comparison, has come under scrutiny, most notably from the National Research Council (2009), the PCAST (2016) report and the American Association for the Advancement of Science (2017), but also from several other sources (Cole, 2006, 2008; Haber & Haber, 2008; Mnookin, 2008; Saks, 2010; Saks & Koehler, 2008). Indeed, experiments have shown that fingerprint examiners are fallible; they make mistakes and are influenced by extraneous information (Busey & Dror, 2011; Busey & Loftus, 2007; Dror, 2017; Dror & Cole, 2010; Dror, Kukucka, Kassin, & Zapf, 2018; Edmond, Tangen, Searston, & Dror, 2014). But issues surrounding the validity of fingerprint examination and other forensic sciences are not the focus of the present paper. Our focus is on the performance of fingerprint examiners *relative* to novices rather than *absolute* performance.

Although fingerprint examiners make errors, they are far more accurate than novices in determining whether two prints were left by the same finger (Tangen et al., 2011). In particular, they demonstrate superior "non-analytic" skills in making accurate judgments about fingerprints quickly and with limited information. For example, they can accurately determine whether two prints match after only viewing them for a fraction of a second, or embedded in visual noise (Thompson & Tangen, 2014). More remarkably, after observing four fingerprints from one hand, fingerprint examiners can discern above chance whether a fifth print belongs to the same hand (Searston & Tangen, 2017a). Less research has been devoted to understanding the more *analytic* capabilities of these experts and their meticulous attention to the detailed features within fingerprints. Filling this gap in our understanding of their expertise is crucial given the prevailing method for comparing fingerprints has been the ACE-V (analysis, comparison, evaluation, verification) method, which involves a slow and deliberative process of comparing features (Ashbaugh, 1999) and could be described as highly "analytic" in nature. On peering over the shoulder of an examiner while they work, for instance, it is common to see them physically markup "minutiae" on a crime scene or "latent" fingerprint as points of comparison for sifting through features in the suspect or "rolled" fingerprint (Ulery, Hicklin, Roberts, & Buscaglia, 2016). These in situ observations of fingerprint examiners' comparison process suggest that the "minutiae" or finer features of fingerprints are critical to the decisions that examiners routinely make.

To determine whether two fingerprints "match" or belong to the same finger, examiners select features they perceive to provide sufficient information for making a decision. However, examiners are sometimes inconsistent in whether they think two prints match if asked at different time points (Dror & Rosenthal, 2008), and there is considerable variability in the way different examiners markup features in fingerprints from one occasion to the next (Dror et al., 2011; Neumann et al., 2007; Ulery, Hicklin, Roberts, & Buscaglia, 2014; Ulery et al., 2016). On the same print, some examiners choose to markup only a few points for comparison, while others decide to markup many. The annotations of experts not only differ substantially in minutia count but also in the particular features that are selected (Ulery et al., 2014). Examiners are also more likely to make a decision (rather than an inconclusive judgment) the more features they markup (Ulery et al., 2014). Disagreement among expert fingerprint examiners about the way features ought to be interpreted may be problematic in courts when evaluating evidence. However, there is little evidence that discrepancies in print annotation necessarily *cause* examiners to be more or less accurate in their discrimination of prints. For instance, it is unclear whether differences in feature markings reflect how examiners see and interpret data or simply how they choose to document their interpretations. Clear and complete fingerprints would seem easier to compare than noisy or distorted prints and therefore provide a greater number of features to identify and markup.

Evidence that examiners are inconsistent in their choice of features when initially analyzing a print also says little about their ability to match prints accurately relative to novices (Tangen et al., 2011). Moreover, inconsistency in the way examiners markup fingerprints provides little insight into their expertise without a novice comparison. Relative to what or to whom are these experts inconsistent? To our knowledge, fingerprint examiners have not been compared to any control group on such feature selection tasks. In the current study, we are therefore primarily interested in understanding how expert examiners differ from novices in their choice of features and the variability in their choices.

Looking to other domains, eye-tracking studies suggest that those with training and experience attend to different features or objects than those without. Experienced radiologists, for instance, spend more time fixating on task-relevant or diagnostic regions relative to non-diagnostic regions and are also quicker to fixate on these critical regions relative to novices (Drew et al., 2013; Krupinski, 1996; Krupinski, Graham, & Weinstein, 2013). Chess experts are far quicker to fixate on relevant parts of the chess board (i.e., the regions associated with the best move) when compared to novices chess players (Sheridan & Reingold, 2014). The search patterns of fingerprint examiners when looking at prints are also far more constrained than those of novices (Roads et al., 2016), typically paying attention to the core (the center of a fingerprint that forms different pattern types) and the delta (a Y-shaped point in a fingerprint where ridges meet) regions (Busey, Nikolov, Yu, Emerick, & Vanderkolk, 2017). Although eye gaze does not necessarily correlate with attention and perception, these studies provide evidence that training and experience can be associated with a greater focus on features understood to be important in accomplishing the task at hand.

Self-report and behavioral measures tell a similar story. Medical diagnosticians have superior memory for X-rays containing abnormalities than X-rays that do not (Myles-Worsley, Johnston, & Simons, 1988). Professional footballers are more sensitive to meaningful, game-relevant changes in football scenes relative to novices, but not scenes that are unrelated to football (Werner & Thies, 2000).

When matching faces, trained forensic face examiners – who are demonstrably better than the general population at matching unfamiliar faces – consider more diagnostic features like the ears, eyes, nose and scars more useful than novices do (Towler, White, & Kemp, 2017). The matching performance of expert fingerprint examiners also drops substantially when more diagnostic areas of the fingerprint are removed (Busey et al., 2017).

The totality of these findings suggests that expertise in many perceptual domains hinges on an ability to distinguish signal from noise (Kellman & Garrigan, 2009). Some features are more fleeting, coming and going as circumstances change. Others are more "sticky"; they can be reliably used to distinguish one category from another. An expert has learned to appreciate these subtleties, becoming more sensitive to the appearance or manifestation of particular features that will provide them with the most useful information, and those they ought to ignore. Moreover, experts ought to converge on these perceptions, because training on the same class of stimuli should calibrate these sensitivities similarly toward the most useful features. If experts differ from novices in which features they deem useful and experts independently converge on a small number of these features, then one clear motivation for future training would be to learn what information is useful and what is not.

An appreciation for what is useful when categorizing natural objects relies heavily on instantiated features (Brooks & Hannah, 2006). Many of us, for example, may describe a bird as a creature that sings and has wings and feathers. However, using solely these descriptors can lead to many errors of classification. To take Brooks' (2005) example, we may incorrectly classify an opera singer with a feather boa on an airplane as a "bird" based on those verbal descriptors. Explicit rules can serve to name objects of perceptual learning and provide instructions, but they do not provide sufficient conditions for identification. For example, a wing of one species of bird can appear very different depending on whether the bird is sitting quietly, flying or fighting with a conspecific, and also its sex or stage of development. Familiarity with different instantiations of a feature influences categorization more so than verbalisable rules (Brooks & Hannah, 2006). Perceptual expertise often hinges on an appreciation for what instantiated features can be most readily used to discriminate between a category object or identity.

Extending this notion to fingerprint examination, a delta may be a useful landmark when analyzing fingerprints. Knowing this rule of thumb may help guide beginners early in training, but this feature – which is easy to spot and describe – would be insufficient for distinguishing prints from one another. Instead, the specific appearance or manifestation of the delta, and how deltas tend to vary between individuals or within fingerprints from the same individual, is critical for perceptual expertise (Kramer, Young, & Burton, 2018). The significance of a delta, for instance, can vary depending on the nature of the impression. There are sometimes multiple deltas on a print, and sometimes none at all. One may have light, curving edges, whereas another may appear more jagged and twisted. Features may also be smudged so significantly that they are no longer helpful for distinguishing a particular fingerprint from other prints. Yet, because of their exposure to the way features tend to appear across thousands of fingerprints, experts are implicitly aware of these nuances. Novices less so.

A richer understanding of how expert fingerprint examiners and novices differ in their selection of features and their variants could yield promising insights about their expertise and how to train it. But our goal is not to generate a list of features that are generally useful or those that are not. Instead, knowing what experts consider most useful and least useful, whether they independently converge on a small number of features, how they compare to novices, and how this all may change depending on the print, provides an insight into their expertise and could inform recruitment and perceptual training in the future. Rather than learning a bank of verbalisable features, it may be more beneficial to efficiently localize and compare important features with an implicit understanding that their usefulness may change depending on the appearance of a print. In fact, work of this kind is already underway with training modules focused on attentional highlighting (Roads et al., 2016).

In the present study, we use a feature annotation paradigm to explore whether novices and experts differ in the features they consider to be most and least useful, and whether experts are more likely to converge on the features they select. We compare the degree to which experts differ from one another, and from novices, in their choice of features by comparing the coordinates of those features within each print. Unlike previous studies of this kind, we are interested only in the features that individuals consider *most useful* and *least useful*, and not the *number* of features they believe to be necessary for making a comparison. Experts may disagree on exactly how many features require further inspection when comparing two prints, and about which features are necessary to make such a judgment, but they could still agree with each other more than novices about which features are most and least useful. We predict that the points experts choose to be most useful will differ significantly in location from the points novices choose to be most useful (Prediction 1) and that the points experts choose to be most useful will have less dispersion than those chosen by novices (Prediction 2). Similarly, we predict that the points experts choose to be least useful will differ significantly in location to the points novices choose to be least useful (Prediction 3) and that the least useful points chosen by experts will show less dispersion than those chosen by novices (Prediction 4).

2 | METHOD

2.1 | Participants

A total of 30 novices (16 females, M_{age} = 30.5, *SD* = 10.2) with no prior fingerprint experience, and 30 experts (11 females, M_{age} = 42.7, *SD* = 5.8) with an average of 11.0 years of experience (*SD* = 7.2) participated in the current study. Novices were recruited primarily from The University of Queensland community and from the broader Brisbane region. Experts were primarily recruited at the Queensland Police Service Fingerprint Bureau but also from the New South Wales Police Force and from the Australia Federal Police. On completing their annotations, the data from each participant were converted into digital form through LiveCode Community 9.0.0, which then output a single plain text file for each participant. The de-identified data for this experiment are available in the "Materials and Experiment" section of this experiment's Open Science Framework pre-registration (https://osf.io/rxe25/).

2.2 | Materials

A set of 100 fingerprint images were obtained from a ground truth database that we developed. We selected 50 "rolled" fingerprints and 50 latent fingerprints, including 10 impressions from each finger type (e.g., left thumb, right middle) from the database. No two fingerprints were from the same source. Each fingerprint image was cropped to a square such that the entire image was filled with ridge detail. This preprocessing step ensured that participants did not mark points on the image outside the fingerprint (further detail can be found in this pre-registration: https://osf.io/gqc9a/). The images were shuffled in six different randomized sequences using the Page Shuffle application (version 10.1; Miln, 2018). The fingerprint images were presented individually on white A4 paper (80 gsm), bound together in a booklet. Participants were also given a red and green pencil to the draw on the images.

2.3 | Procedure and design

Each of the experts and novices were given a booklet containing the same 100 fingerprints (one per page). They were asked to pencil one small green circle at the center of an area or feature they thought was most useful for distinguishing that fingerprint from other fingerprints and one small red circle at the center of an area or feature they thought was least useful for distinguishing that fingerprint from other fingerprints. We also gave participants examples to clarify what we wanted from them. For example, if asked what one might select when trying to distinguish an ibis from other birds, they may see its long slender beak. We also included an example print that had been marked with a green and red dot to demonstrate the size we expected the circles to be. The full instructions can be found on the OSF preregistration (https://osf.io/gqc9a/). It may be that when comparing two prints side-by-side, fingerprint examiners perceive different features to be useful on a case by case basis depending on the prints at hand. However, here we are interested in the features that participants consider generally useful in the absence of any single comparison print.

2.4 | Data analysis

To gather the coordinates of all the most useful and least useful points for all the images across every participant, we overlaid a 50×50 transparent grid over each image. We entered the locations of the points manually into our LiveCode program. When a circle spanned several grid squares, we used the square that was covered the most or the one closest to the circle's center. Given that we obtained both an x and y coordinate for each point, the data we gathered were multivariate. It was also distributed non-normally, because the points were sometimes dispersed across several regions of an image rather than evenly around a central mean. Therefore, to test for differences between experts and novices in the points they chose, we decided to use a permutational multivariate analysis of variance (PERMANOVA; Anderson, 2005). PERMANOVA is a non-parametric test that calculates a Fisher's F-statistic on the basis of permutations – where the data are shuffled and resampled many times – and Euclidean distances between points instead of the differences between two normal distributions. This test mirrors the classic ANOVA in many ways but allows for more rigorous and meaningful analysis of high-dimensional systems, even ones distributed non-normally. PERMANOVA can discern if the centroid of the points chosen by experts differs significantly from the centroid of the points chosen by novices. That is, the analysis will indicate whether the locations of the expert points differ in location from the novice points.

In addition to testing for differences in location, we also assessed the dispersion (or agreement) of each group's points using a non-parametric test for homogeneity of multivariate dispersions (PERMDISP; Anderson, 2004) that also calculates significance via permutations. We then determined the direction of any group difference by observing the mean dispersion of each group around a centroid. These analyses were conducted for the most useful and least useful points as chosen by each group for every image in our set. All of the data analysis can be found in the "Analysis" section of the OSF preregistration (https://osf.io/w2tsa/).

3 | RESULTS

Before discussing results, we will first provide an illustration of the relatively unconventional analyses that we used. Although we provide only a few illustrative examples of the analyses here, the full data analysis workflow and data visualization for each print is available on the Open Science Framework (https://osf.io/w2tsa/). Figure 1 presents example spatial representations of the points that experts and novices chose as most useful. The heatmaps detail the choices of experts and novices overlaid onto their respective prints. The plots beside them represent the same points and their group centroids, as well as the dispersion of each group's points around the centroid. The top row of plots depicts an instance where the permutational analysis of variance (PERMANOVA) and permutational test for dispersions (PERMDISP) were both significant – where expert points differed in location and dispersion from the novice points. The bottom row depicts a case where the groups differed neither in location nor dispersion.

For every image in our set, we conducted a permutational analysis of variance and permutational test of dispersions like those presented in Figure 1 to compare experts to novices in the points they chose as most useful and least useful. We then computed the percentage of fingerprints where the groups differed significantly and generated a distribution of the F values obtained from running the analyses on each print. Figure 2 displays the distribution of these F values in

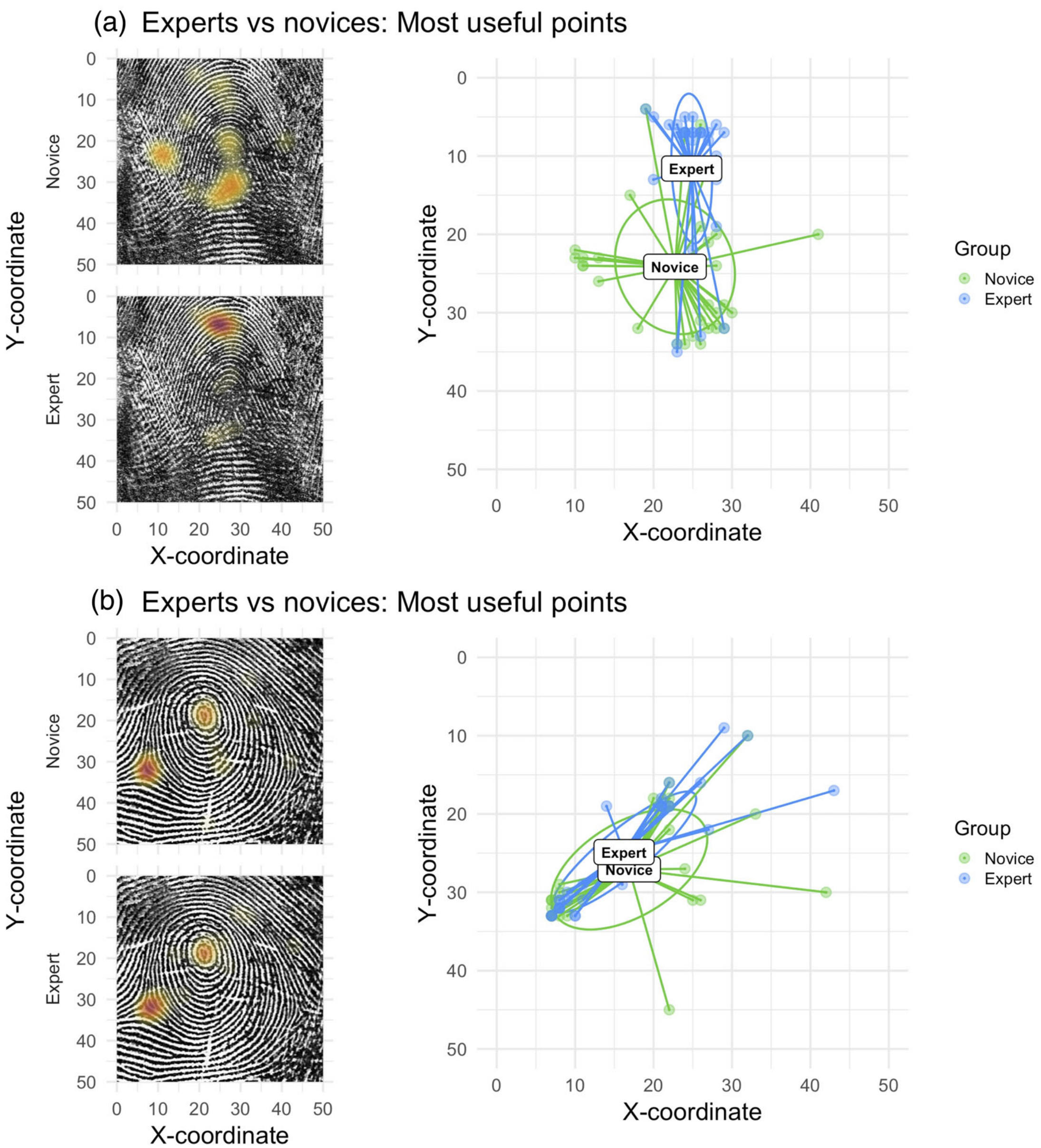

FIGURE 1 A visualization of group differences in their choice of the most useful feature. (a) shows an example image where the choices of experts and novices differed in location on the print displayed and where expert points were more tightly clustered. (b) illustrates a case where experts and novices differed neither in location nor dispersion. The fingerprints on the left include heat maps of the points chosen by each group. The plots on the right spatially display the centroid and variances (1 SD ellipse) of each group around the respective group centroid [Colour figure can be viewed at wileyonlinelibrary.com]

ascending order across of the four types of analysis: differences in location and differences in dispersion for the most useful points and least useful points. For clarity, the fingerprints where we found less dispersion among novice points were coded negatively.

3.1 | Most useful points

One-way permutational analyses of variance revealed that the locations of the most useful points that experts chose differed significantly from the points that novices chose on 51% of images. One-way permutational tests of dispersion revealed that the distributions of each group's points differed significantly on 48% of the fingerprints. The dispersion of the expert points was lower than the novice points on 47 of the 48 images where differences were found. Figure 1 depicts two example cases to illustrate the variability of results. It shows an example case, where expert and novice points differed in location and dispersion and another case where the groups differed in neither.

3.2 | Least useful points

One-way permutational analyses of variance revealed that the locations of the least useful points that experts chose differed significantly from

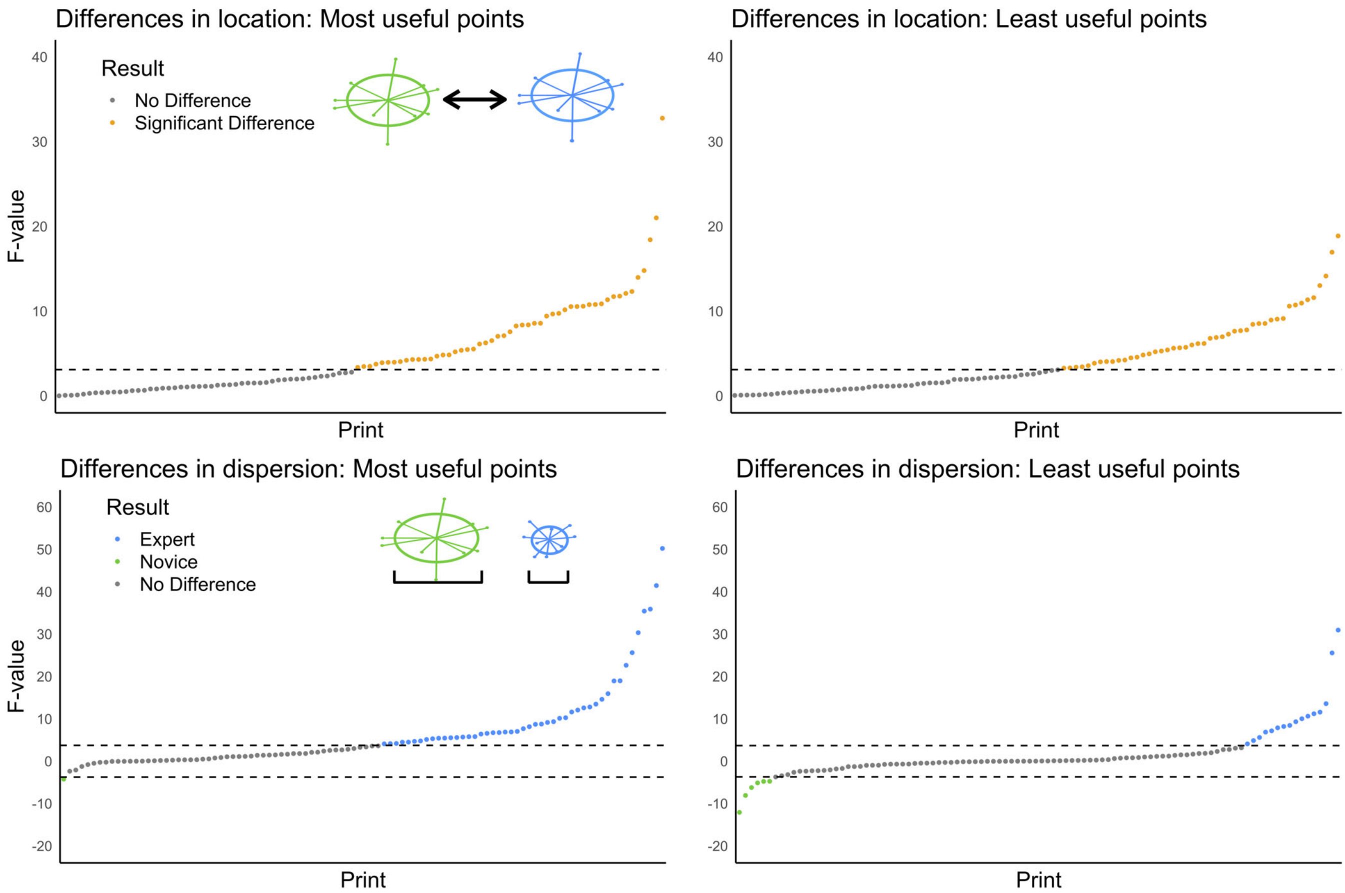

FIGURE 2 An overview of all analyses across the set of fingerprints. The distributions for the useful points are displayed on the left and unuseful points on the right. On top, the Fisher's F-values from the permutational multivariate analyses of variance (PERMANOVA) between experts and novices across the set of prints are displayed in ascending order. Significant *F* values are coded in gold and non-significant *F* values in gray. The dashed line indicates the threshold for significance. The plots on the bottom display the average distance of each group's points from their centroid, organized in ascending order according to *F* values from the permutational tests of homogeneity of multivariate dispersions (PERMDISP). Blue points denote instances where experts showed less dispersion whereas points in green signify when novices showed less dispersion, which are also coded negatively. The dashed lines represent thresholds for significance; cases where novices show more agreement fall below the bottom line, cases with more expert agreement fall above the top line and cases where groups did not differ fall between the dashed lines [Colour figure can be viewed at wileyonlinelibrary.com]

the points that novices chose on 46% of the images. One-way permutational tests of dispersion revealed that the distributions of each group's points differed significantly on 22% of fingerprints and that the dispersion of the expert points was lower than the novice points on 16 of the 22 images where differences were found. In Figure 3, we provide a visualization of the data. Depicted below is an example print where expert and novice least useful points differed in location, and where expert points were less dispersed. We also show an example print where the groups differed in neither dispersion nor location.

3.3 | Exploratory analyses

3.3.1 | Aggregating across groups

In addition to the planned analyses above, we conducted exploratory analyses to determine whether there were differences in location and dispersion between the most useful and least useful points, regardless of which group chose them. One-way permutational analyses of variance revealed that the locations of the most useful points differed significantly from the least useful points on 88% of the images. That is, useful points were frequently located in different regions to less useful points. One-way permutational tests for dispersions revealed that the variances between the types of points differed significantly on 87% of the images. The dispersion of the more useful points was lower than less useful points in every case. That is, the less useful points were never clustered more tightly than the more useful points.

3.3.2 | Examiner commentary

We approached an expert fingerprint examiner who had not participated in the experiment to view each of the 100 expert and novice

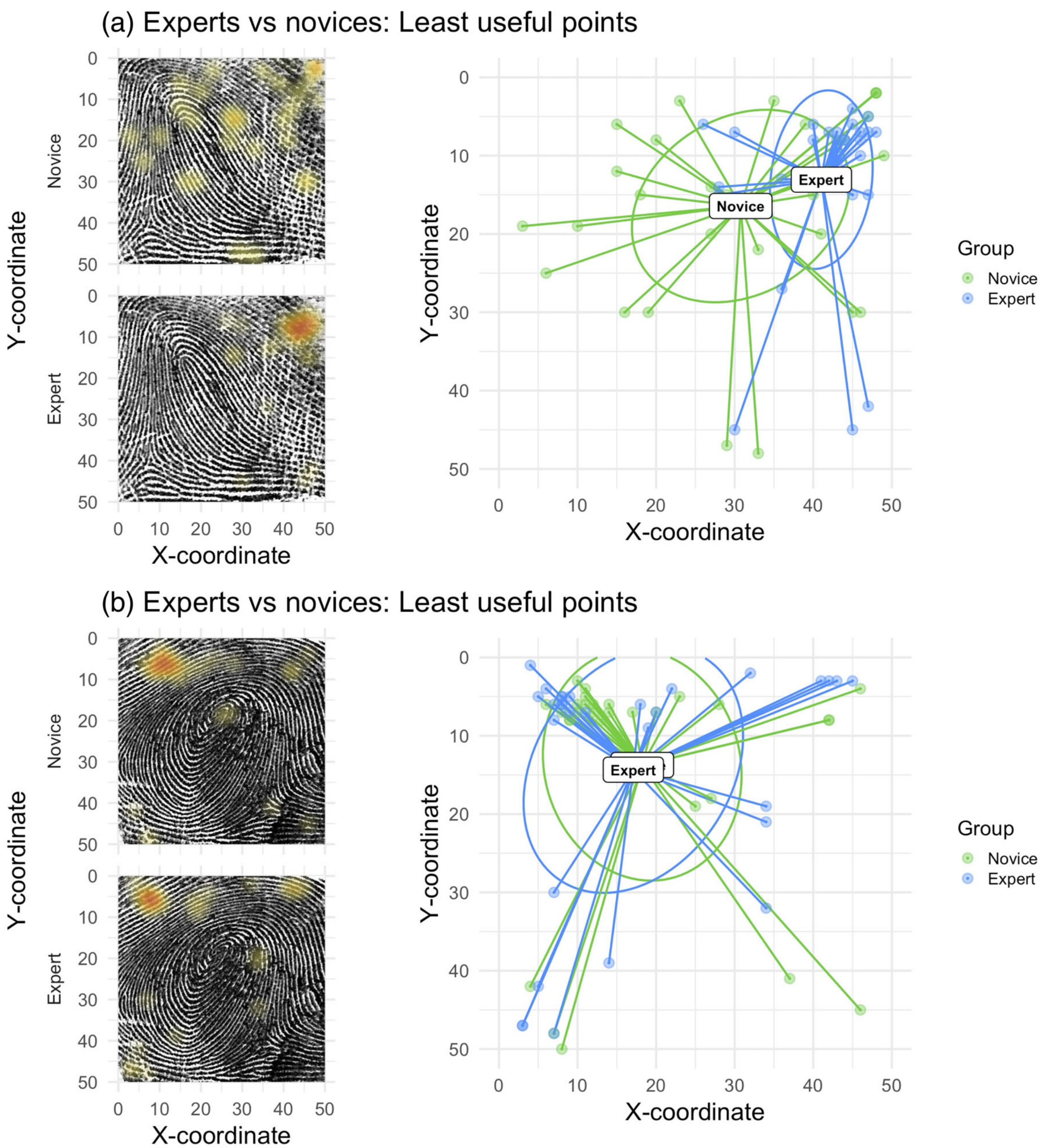

FIGURE 3 A visualization of group differences in unuseful feature choice. (a) shows an example fingerprint where the feature choices of experts and novices differed in location, and where expert points were more tightly clustered. (b) illustrates a case where experts and novices differed neither in location nor dispersion. The fingerprints on the left are overlaid with heat maps of the points chosen by each group. The plots on the right spatially display the centroid and variances (1 SD ellipse) of each group around the respective group centroid [Colour figure can be viewed at wileyonlinelibrary.com]

useful feature heat maps. For the prints where the groups differed, both experts and novices often picked out the core and the delta. However, the novice features were more often described as being more "distributed "or "scattered" as well as "unclear," "not distinctive" or "non-selective." Novices were also more likely to choose features such as "creases," "white lines," "scars," "spots" and "smudges." These features are either imperfections that occur when a fingerprint is laid, impermanent folds in the skin, or features generally not classified as ridge detail. The features in the expert heat maps were often described as "ridge detail" that is "clear," "distinct" or "visible," and were frequently given labels of small but distinct minutiae such as "short ridges," "lakes," and "bifurcations." After viewing each print, the examiner retrospectively commented that novices were more variable in their feature choices and the features themselves were "not necessarily best as a target group for starting a comparison in terms of clarity, ridge detail and being able to locate them quickly and easily" in another print. He further added that they tended to opt for "smudges, creases and other material that was not ridge detail." Features chosen by experts, on the other hand, were described as "clear and locatable" ridge detail.

4 | DISCUSSION

The aim of the present study was to investigate whether experts differ from novices in their choice of the most useful and least useful features, using fingerprint examination as a test bed domain. We were also interested in whether experts are more consistent in their feature choices. The findings partially support our first prediction: the location of the most useful features that experts chose differed significantly from those chosen by novices on half of the images. These findings suggest that experts and novices often consider different features useful when making perceptual decisions in the domain of fingerprints. A likely explanation for why these groups differed in their feature choices so frequently is that the considerable training examiners have undertaken has equipped them with an appreciation for what features will provide them with information capable of supporting a decision. However, we collected data only from novice and expert examiners, not trainees. One next step could be to investigate how perceptions of usefulness change throughout training by examining differences between trainees with varying levels of experience.

We also cannot be certain that the choices made by experts are necessarily "better," because we do not yet know what features aid or impede one's ability to discriminate between prints, but such an interpretation does converge with the broader literature on perceptual expertise. For example, those with greater experience in domains like chess and medicine selectively attend to the more diagnostic or relevant features of stimuli from their domain of expertise (Krupinski, 1996; Krupinski et al., 2013; Roads et al., 2016; Sheridan & Reingold, 2014). Experts were considerably more consistent in their feature choices than novices across the set. The features experts chose were less dispersed than novices on about half of the prints, while showing more dispersion on only a single print. Standardization of training practices and feedback throughout training may explain why choices are more consistent for experts than for novices as each examiner gains a similar appreciation for what is useful.

Our third prediction was also partially supported: the features experts chose as less useful differed significantly in location from those that novices chose on almost half of the images. We also found evidence to support our fourth prediction: experts were likely to agree more on which features were least useful compared to novices but only on a small number of the fingerprints. Further exploratory analyses showed that the most useful and least useful features selected were consistently located in different regions of the fingerprint, irrespective of who chose them, and the more useful features selected consistently clustered closer together relative to less useful features.

An appreciation for what is more useful necessarily sensitizes examiners to what is less useful. Years of marking up, analyzing and comparing fingerprints possibly explains why experts see different features as less useful compared to novices. But although experts tended to agree more than novices in their choice of less useful features, these differences seldom occurred. One possible reason for why we did not find more consistent agreement among experts in their selection of less useful features is that they tend to be located at the edges of fingerprints, both in our experiment but also in others (Busey et al., 2017; Roads et al., 2016). Moreover, anything not considered useful might be regarded as useless, so participants have more to choose from when selecting the least useful feature. Less useful features are therefore highly variable by their nature, and this large variation can make group differences more difficult to detect.

Previous research has shown that experts can be inconsistent with themselves and with other experts in the number of features they choose for comparison (Dror et al., 2011; Neumann et al., 2007; Ulery et al., 2014, 2016). Despite this inconsistency within experts, our findings demonstrate that experts are more consistent than novices. That is, experts may disagree on exactly how many features are needed to decide whether two prints match, but they agree on what features are highly informative to a greater extent than novices. It would seem that experience with fingerprints changes one's perception of the relative importance or usefulness of different features.

Of course, experts and novices did not always differ in their feature choices, nor in how widely their choices varied, and there are several possible explanations for why we did not find expert–novice differences more frequently. The first is that the methods we used to analyze the data were fairly coarse. On many of the fingerprint images, there may have been several features regarded as highly useful and several regarded as less useful, but measured only variation in Euclidean distance around each group's centroid. These analyses were not well equipped to sufficiently detect expert–novice differences in the presence of *multiple clusters* of highly useful or useless features. Low sensitivity to detect multiple clusters of feature choices may in part explain the lack of consistency in the findings. It could also be that verbal descriptors may be a more appropriate measure of expert–novice differences in feature choice. However, the features participants choose may not necessarily be verbalisable, nor does training to search for and compare features hinge on an ability to label what one sees.

Another explanation for why the groups differed inconsistently is that experts rely heavily on prior experience to inform their decisions and a deliberative feature selection task like ours may not reflect this kind of intuitive expertise. Fingerprint examiners are indeed better than novices at making fast and intuitive judgements based on a whole set of features distributed across fingerprints, even without corresponding details (Busey & Vanderkolk, 2005; Searston & Tangen, 2017a; Thompson & Tangen, 2014). But the presence of such non-analytic abilities does not preclude experts from having superior analytic abilities too. Fingerprint examiners report relying on a meticulous feature analytic process of comparison (Ashbaugh, 1999; Ulery et al., 2014, 2016). Consistent with such accounts, our findings demonstrate that experts do indeed appreciate different features to novices; and they are more consistent in the features they perceive to be useful. Future research ought to investigate whether the features that experts and novices subjectively regard as useful are in fact differentially more helpful for discriminating fingerprints.

It may also be that experts and novices rely on very similar features when making perceptual decisions. In our exploratory analyses, we found far more robust differences in location and dispersion when we compared the more useful and less useful points, aggregating across groups. In other words, the difference between the more useful features and less useful features was far greater and more consistent than any expert-novice difference. The features that are most salient may often be the most useful. Two focal points on many fingerprints are features known as "cores" and "deltas" (see fingerprint B in Figure 1, where the core and delta are highlighted on the heatmap). Much of the time, both groups may consider highly salient features like the core and delta to be "useful" or characteristic of any individual print, but experts may glean far more information from the same ridge detail than novices. Indeed, eye-tracking research has demonstrated that although experts are more constrained in where they look within fingerprints, there is overlap in the areas that experts and novices attend to (Roads et al., 2016). Future research could reveal the cognitive processes underlying feature selection and attention, and how these processes change with training; however, this latter explanation may account for why experts and novices did not differ in their selection of features on many prints. In any case, inconsistency in group differences suggests a moderating effect; some conditions may call for greater expert knowledge whereas others may not.

We think a plausible moderator of expert-novice differences in feature choice is feature clarity in light of the commentary provided by the examiner we consulted. Clarity is an important determinant for selecting fingerprint features (Ulery et al., 2014, 2016) and consistency in feature choice among examiners is far higher in clear areas than noisy areas (Ulery et al., 2016). Novices may focus on salient features (like the core and delta), and these same salient features may also be most informative to experts if they are clear and undistorted. The visibility of features like the core or delta might make comparison much easier: if the salient features do not match then no further examination is required, so a comparison can be made quickly (Kellman et al., 2014). In circumstances of high clarity, it would therefore be unsurprising to see little difference between each group's choice in features but if features are degraded or distorted or even entirely absent, the groups may then start to diverge in what they attend to. Novices may persist in attending to features even if they are not sufficiently clear to be useful when making a comparison, or they may be tempted to choose features that have surface-level saliency but are often unreliable, such as creases and smudges. Experts, in the knowledge that they can only obtain useful information from clear ridge detail, will go on to search in other clearer regions for seemingly less obvious features such as lakes and bifurcations.

Further evidence for this explanation can be observed in Figure 1. Whereas both groups consider the core and delta to be useful in the clear, undistorted bottom fingerprint, their choices have little correspondence in the top fingerprint. The core and delta of the top print are heavily distorted, but novices have persisted in selecting them as useful. Experts, by contrast, have shifted their attention to clearer regions containing a collection of smaller features.

Prioritizing clarity may also explain inconsistent expert-novice differences in their choice of less useful features. Experts may consider highly obscured features of a fingerprint to be least useful. Novices on the other hand may persevere with choosing clearer but less-salient ridge detail. Indeed, in Figure 3 (top), experts have consistently chosen a single highly distorted region in the top right corner of the print, whereas novices have chosen several clearer regions around the edge of the print. It is plausible then that skilled feature choice becomes more noticeable only in more challenging circumstances – when details are less clear and more ambiguous. Consistent with this interpretation, Busey and Parada (2010) suggest on the basis of eye-tracking research that fingerprint examiners possess an implicit understanding of what regions contain the most useful information but in cases of high noise, examiners will inspect different regions to identify informative features.

The recognition-primed decision (RPD) model (Klein, 2008) is a theoretical framework that can shed further light on this latter interpretation of when and why experts differ from novices in their choice of features. The RPD model describes how people use a repertoire of patterns that they have gained through considerable experience to make decisions quickly and accurately (Klein, Calderwood, & Clinton-Cirocco, 1986). These patterns signal relevant cues, provide expectancies, and identify plausible goals. When people need to make a decision quickly, they match the situation to the patterns that they have learned. If they find a clear match, they can then evaluate if the course of action is viable by mentally simulating it and seeing if it plays out as desired. If it does, then they can act accordingly. If not, they can shift their attention to the next typical plan of action. Hierarchical option-generation strategies like the RPD model have been used to describe decision-making in domains such as fireground command (Klein et al., 1986), chess (Klein, Wolf, Militello, & Zsambok, 1995) and sport (Raab & Johnson, 2007). In the case of fingerprints, when examiners come across a print, they will most likely attend to the core and the delta. If these areas are smudged or blurred, they may no longer be viable sources of information and experienced examiners know they must find other features to attain the information they need. Conversely, novices have no implicit knowledge of the common features in fingerprints, so their mental simulations are not rich enough to encourage them to look elsewhere. Differences between experts and novices in feature choice may therefore emerge only in ambiguous circumstances, as feature choice begins to align less and less with saliency and more with feature clarity. An avenue for future research is to investigate how clarity and distortion affect how experts and novices choose and interpret features. Better understanding the role of clarity and distortion could inform how novices are trained to discriminate between categories and identities in perceptual domains like fingerprint examination. A core component of acquiring expertise may be an appreciation for the utility of different features and how utility can vary from one impression or instantiation to the next.

Given adequate perceptual training, humans can excel at making decisions in ambiguous situations. An ability to change what features one attends to in more challenging situations may be why humans still outperform computerized systems in many perceptual domains, including fingerprint examination. Fingerprint examiners can make accurate decisions even when prints are clouded in heavy noise (Thompson & Tangen, 2014). For latent print comparisons, the capabilities of computerized systems are yet to meet those of human experts (Dror & Mnookin, 2010). Latent fingerprints, often collected at crime scenes for instance, are often partial, noisy, distorted, smudged, and typically contain less information than fingerprints collected under more controlled conditions. Stimulus features like these make the task of comparing a latent fingerprint to a rolled fingerprint significantly more complex (Kellman et al., 2014).

The present study provides insight into how experts and novices select features to make discrimination judgements, in the context of fingerprint examination. Using a simple paper and pencil experiment, we asked experts and novices to mark features in fingerprints they considered to be most and least useful for distinguishing the print. We measured the extent to which they agreed on their feature choices. Sometimes experts focused on different fingerprint features to novices, but not always, and most of the time experts agreed more than novices, but not always. We think the inconsistency in expert–novice differences arises from the varied clarity of the features across the set of fingerprints. Obscured and distorted features may mean that useful information is more difficult to find, and experts may excel more than novices when working under such conditions. Future research could investigate the veracity of this hypothesis, but it may be that flexibly shifting one's attention depending on the information available is a critical skill that sets experts apart from novices.

ACKNOWLEDGEMENTS

The authors thank the Australian fingerprint examiners who participate in our research for giving their time and expertise so generously. This research was supported by grant no. LP170100086 from the Australian Research Council (ARC) to Tangen, Searston and Edmond.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The reported study was cleared in accordance with the ethical review processes of The University of Queensland and within the guidelines of the National Statement on Ethical Conduct in Human Research. The research was conducted in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

In line with the CRediT taxonomy, S.G.R. contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualization, writing and editing. J.M.T. contributed to conceptualization, methodology, software and reviewing and editing. R.A.S. and G.E. contributed to reviewing and editing, D.J.M. contributed to investigation.

DATA AVAILABILITY STATEMENT

The data and code for each individual novice and expert participant used to produce our results and plots are available here, with the exception of participants' age and years of experience.

ORCID

Samuel G. Robson https://orcid.org/0000-0002-2777-9347

REFERENCES

American Association for the Advancement of Science (AAAS) (2017), Forensic Science Assessments: A Quality and Gap Analysis-Latent Fingerprint Examination, *(Report prepared by William Thompson*, John Black, Anil Jain, and Joseph Kadane), https://doi.org/10.1126/srhrl.aag2874

Anderson, M. J. (2004). *PERMDISP: A FORTRAN computer program for permutational analysis of multivariate dispersions (for any two-factor ANOVA design) using permutation tests*. Auckland, New Zealand: Department of Statistics, University of Auckland.

Anderson, M. J. (2005). *PERMANOVA: A FORTRAN computer program for permutational multivariate analysis of variance*. Auckland, New Zealand: Department of Statistics, University of Auckland. Retrieved from https://doi.org/10.1002/9781118445112.stat07841

Ashbaugh, D. R. (1999). *Quantitative-qualitative friction ridge analysis: An introduction to basic and advanced ridgeology*. Boca Raton, FL: CRC press, an imprint of Taylor & Francis Group.

Brooks, L. R. (2005). The blossoms and the weeds. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *59*(1), 62–74. Retrieved from https://doi.org/10.1037/h0087462

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules.". *Journal of Experimental Psychology: General*, *135*(2), 133–151. https://doi.org/10.1037/0096-3445.135.2.133

Busey, T. A., & Dror, I. E. (2011). Chapter 15: Special abilities and vulnerabilities in forensic expertise. In E. H. Holder, L. O. Robinson & J. H. Laub (Eds). The fingerprint Sourcebook (pp. 1–23) Washington, DC: NIJ Press.. Retrieved from https://www.ncjrs.gov/pdffiles1/nij/225320.pdf

Busey, T. A., & Loftus, G. R. (2007). Cognitive science and the law. *Trends in Cognitive Sciences*, *11*(3), 111–117. https://doi.org/10.1016/j.tics.2006.12.004

Busey, T. A., Nikolov, D., Yu, C., Emerick, B., & Vanderkolk, J. (2017). Characterizing human expertise using computational metrics of feature diagnosticity in a pattern matching task. *Cognitive Science*, *41*(7), 1716–1759. https://doi.org/10.1111/cogs.12452

Busey, T. A., & Parada, F. J. (2010). The nature of expertise in fingerprint examiners. *Psychonomic Bulletin and Review*, *17*(2), 155–160. https://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*(4), 431–448. https://doi.org/10.1016/j.visres.2004.08.021

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

Cole, S. A. (2006). "Implicit testing": Can casework validate forensic techniques? *Jurimetrics*, *46*, 117–128.

Cole, S. A. (2008). The 'Opinionization' of fingerprint evidence. *BioSocieties*, *3*(1), 105–113. https://doi.org/10.1017/S1745855208006030

Drew, T., Vo, M. L.-H., Olwal, A., Jacobson, F., Seltzer, S. E., & Wolfe, J. M. (2013). Scanners and drillers: Characterizing expert visual search through volumetric images. *Journal of Vision*, *13*(10), 3–3. https://doi.org/10.1167/13.10.3

Dror, I. E. (2017). Human expert performance in forensic decision making: Seven different sources of bias†. *Australian Journal of Forensic Sciences*, *49*(5), 541–547. https://doi.org/10.1080/00450618.2017.1281348

Dror, I. E., Champod, C., Langenburg, G., Charlton, D., Hunt, H., & Rosenthal, R. (2011). Cognitive issues in fingerprint analysis: Inter- and intra-expert consistency and the effect of a "target" comparison.

Forensic Science International, 208(1–3), 10–17. https://doi.org/10.1016/j.forsciint.2010.10.013

Dror, I. E., & Cole, S. A. (2010). The vision in "blind" justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin and Review*, *17*(2), 161–167. https://doi.org/10.3758/PBR.17.2.161

Dror, I. E., Kukucka, J., Kassin, S. M., & Zapf, P. A. (2018). When expert decision making goes wrong: Consensus, bias, the role of experts, and accuracy. *Journal of Applied Research in Memory and Cognition*, *71*(1), 162–163. https://doi.org/10.1016/j.jarmac.2018.01.007

Dror, I. E., & Mnookin, J. (2010). The use of technology in human expert domains: Challenges and risks arising from the use of automated fingerprint identification systems in forensic science. *Law, Probability and Risk*, *9*, 47–67. https://doi.org/10.1093/lpr/mgp031

Dror, I. E., & Rosenthal, R. (2008). Meta-analytically quantifying the reliability and biasability of forensic experts. *Journal of Forensic Sciences*, *53*(4), 900–903. https://doi.org/10.1111/j.1556-4029.2008.00762.x

Edmond, G., Tangen, J. M., Searston, R. A., & Dror, I. E. (2014). Contextual bias and cross-contamination in the forensic sciences: The corrosive implications for investigations, plea bargains, trials and appeals. *Law, Probability and Risk*, *14*(1), 1–25. https://doi.org/10.1093/lpr/mgu018

Haber, L., & Haber, R. N. (2008). Scientific validation of fingerprint evidence under Daubert. *Law, Probability and Risk*, *7*, 87–109. https://doi.org/10.1093/lpr/mgm020

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

Kellman, P. J., Mnookin, J. L., Erlikhman, G., Garrigan, P., Ghose, T., Mettler, E., … Dror, I. E. (2014). Forensic comparison and matching of fingerprints: Using quantitative image measures for estimating error rates through understanding and predicting difficulty. *PLoS ONE*, *9*(5), e94617. https://doi.org/10.1371/journal.pone.0094617

Klein, G. A. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986, September). *Rapid decision making on the fire ground*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting (Vol. 30, No. 6, pp. 576-580). Los Angeles, CA: Sage Publications. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a199492.pdf

Klein, G. A., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes*, *62*(1), 63–69. https://doi.org/10.1006/obhd.1995.1031

Kramer, R. S. S., Young, A. W., & Burton, A. M. (2018). Understanding face familiarity. *Cognition*, *172*, 46–58. https://doi.org/10.1016/j.cognition.2017.12.005

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology*, *3*(2), 137–144. https://doi.org/10.1016/S1076-6332(05)80381-2

Krupinski, E. A., Graham, A. R., & Weinstein, R. S. (2013). Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Human Pathology*, *44*(3), 357–364. https://doi.org/10.1016/j.humpath.2012.05.024

Miln, G. (2018). *Page shuffle (Version 1.0.1)*. [Mobile application software]. Retrieved from https://miln.eu/pageshuffle/

Mnookin, J. L. (2008). Of black boxes, instruments, and experts: Testing the validity of forensic science. *Episteme*, *5*(3), 343–358. https://doi.org/10.3366/e1742360008000440

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 553–557. https://doi.org/10.1037/0278-7393.14.3.553

National Research Council (2009). Strengthening forensic science in the United States: A path forward. Washington, DC: The National Academies Press. https://doi.org/10.17226/12589

Neumann, C., Champod, C., Puch-Solis, R., Egli, N., Anthonioz, A., & Bromage-Griffiths, A. (2007). Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiæ. *Journal of Forensic Sciences*, *52*(1), 54–64. https://doi.org/10.1111/j.1556-4029.2006.00327.x

President's Council of Advisors on Science and Technology (US). (2016). *Report to the president: Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods*. Executive Office of the President of the United States, President's Council of Advisors on Science and Technolgy.

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied*, *13*(3), 158–170. https://doi.org/10.1037/1076-898X.13.3.158

Roads, B., Mozer, M. C., & Busey, T. A. (2016). Using highlighting to train attentional expertise. *PLoS ONE*, *11*(1), 1–24. https://doi.org/10.1371/journal.pone.0146266

Saks, M. J. (2010). Forensic identification: From a faith-based "Science" to a scientific science. *Forensic Science International*, *201*(1–3), 14–17. https://doi.org/10.1016/j.forsciint.2010.03.014

Saks, M. J., & Koehler, J. J. (2008). The individualization fallacy in forensic science evidence. *Vanderbilt Law Review*, *61*, 199–219.

Seamster, T. L., Redding, R. E., Cannon, J. R., Ryder, J. M., & Purcell, J. A. (1993). Cognitive task analysis of expertise in air traffic control. *The International Journal of Aviation Psychology*, *3*(4), 257–283. https://doi.org/10.1207/s15327108ijap0304_2

Searston, R. A., & Tangen, J. M. (2017a). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin and Review*, *24*(4), 1324–1329. https://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*, *12*(6), 1–14. https://doi.org/10.1371/journal.pone.0178403

Sheridan, H., & Reingold, E. M. (2014). Expert versus novice differences in the detection of relevant information during a chess game: Evidence from eye movements. *Frontiers in Psychology*, *5*, 1–6. https://doi.org/10.3389/fpsyg.2014.00941

Sowden, P. T., Davies, I. R. L., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 379–390. https://doi.org/10.1037/0096-1523.26.1.379

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

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*, 737. https://doi.org/10.3389/fpsyg.2014.00737

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. https://doi.org/10.1037/xap0000108

Ulery, B. T., Hicklin, R. A., Roberts, M. A., & Buscaglia, J. A. (2014). Measuring what latent fingerprint examiners consider sufficient information for individualization determinations. *PLoS ONE*, *9*(11), e110179. https://doi.org/10.1371/journal.pone.0110179

Ulery, B. T., Hicklin, R. A., Roberts, M. A., & Buscaglia, J. A. (2016). Data on the interexaminer variation of minutia markup on latent fingerprints. *Forensic Science International*, *264*, 158–190. https://doi.org/10.1016/j.dib.2016.04.068

Werner, S., & Thies, B. (2000). Is "change blindness" attenuated by domain-specific expertise? An expert-novices comparison of change detection in football images. *Visual Cognition*, *7*(1–3), 163–173. https://doi.org/10.1080/135062800394748

How to cite this article: Robson SG, Searston RA, Edmond G, McCarthy DJ, Tangen JM. An expert–novice comparison of feature choice. *Appl Cognit Psychol*. 2020;34:984–995. https://doi.org/10.1002/acp.3676