THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (5), 1023–1040

Psychology Press
Taylor & Francis Group

On the preliminary psychophysics of fingerprint identification

John R. Vokey
University of Lethbridge, Lethbridge, Alberta, Canada

Jason M. Tangen
The University of Queensland, St Lucia, Queensland, Australia

Simon A. Cole
University of California, Irvine, Irvine, CA, USA

For a century, the matching of images of fingerprints has been used for forensic identification. Despite that history, there have been no published, peer-reviewed studies directly examining the extent to which people can correctly match fingerprints to one another. The results of three experiments using naïve undergraduates to match images of fingerprints are reported. The results demonstrate that people can identify fingerprints quite well, and that matching accuracy can vary as a function of both source finger type and image similarity.

Keywords: Fingerprints; Finger print examiners (FPEs); Identification; Matching; Discrimination.

Any reader of crime fiction novels or observer of myriad television police shows would readily accept the proposition that if any system of forensic identification merited the title of established science and unquestionable forensic evidence in legal proceedings it would be fingerprint identification—the matching of latent, print images to finger sources. Indeed, the various CSI ("Crime Scene Investigation") American television series suggest that crime-scene fingerprints are identified in seconds by computer algorithms searching large databases. This representation is

Correspondence should be addressed to John R. Vokey, Microcognition Laboratory, Behaviour and Evolution Research Group, Department of Psychology, University of Lethbridge, Lethbridge, Alberta, Canada T1K 3M4. E-mail: vokey@uleth.ca

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant to J.R.V., an NSERC postdoctoral fellowship at McMaster University and a University of Queensland Early Career Researcher grant to J.M.T. This material is partially based upon work supported by the National Science Foundation under Grants SES-0115305 and IIS-0527729 and by the National Institutes of Health under Grant HG-03302. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author (S.A.C.) and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

We thank Jeff Boychuk, Stan Govenlock, Erica Jeffery, Megan Torry, Jillian Bailey, and Sarah Spence for their assistance in collecting the data. Portions of this work were presented at the Canadian Society for Brain, Behaviour and Cognitive Science in 2004, and the Banff Annual Seminar in Cognitive Science in 2004 and 2005.

© 2008 The Experimental Psychology Society

http://www.psypress.com/qjep
DOI:10.1080/17470210802372987

misleading; although computer algorithms are increasingly being relied upon to match "ten-prints" (sets of 10 intentionally generated fingerprint impressions; Cherry & Imwinkelried, 2006), there is no computer algorithm that is yet relied upon for the purportedly more difficult task of identifying the source of lower quality "latent" (or crime-scene) prints. What little data there are suggest that were we to rely on existing computer algorithms to identify latent prints, they would make significant numbers of erroneous attributions (Cappelli, Maio, Maltoni, Wayman, & Jain, 2006; Maio, Maltoni, Cappelli, Wayman, & Jain, 2002; Pankanti, Prabhakar, & Jain, 2002; Wilson et al., 2004), although in fairness it should be noted that existing algorithms were designed as search aids rather than as decision-making tools.

Instead, in real life, latent print attributions are entrusted to human fingerprint examiners or experts (FPEs). Although it is widely assumed that the human experts are more accurate than computer algorithms, there is, in fact, no actual evidence of this claim. Indeed, there is no good measurement of the accuracy of human latent print examiners at all. It is not just that the available evidence is equivocal, but rather that there really has been no peer-reviewed, published, scientific investigation of this very question (e.g., Cole, 2001, 2003, 2005a, 2005b, 2006, 2007; Epstein, 2002; Faigman, 2002; Faigman, Kaye, Saks, & Sanders, 2002; Haber & Haber, 2003; Kennedy, 2003; Loftus & Cole, 2004; Siegel et al., 2006; Specter, 2002; Starrs, 1999; Zabell, 2005). Wertheim, Langenburg, and Moenssens (2006b) provide a recent attempt at assessing FPE accuracy using performances during FPE training exercises. Unfortunately, it is quite flawed. Among other things, it lacked distractor test prints, so false positives could not be assessed. See Haber and Haber (2006) for details, and Wertheim, Langenburg, and Moenssens (2006a) for the authors' response.

Interestingly, at least one pioneer of fingerprinting believed that experts would quickly become unnecessary and that lay juries would eventually evaluate fingerprint evidence (Galton, 1893). Instead, police identification bureau clerks asserted that their experience observing patent (inked) prints and classifying them into categories based on pattern types furnished them with expertise in identifying the source of latent (crime-scene) prints (Cole, 1998). Courts have accepted this claim without any empirical evidence in support of it (Cole, 2004). Indeed, when challenged to produce such evidence, fingerprint examiners and the prosecutors who use the evidence in court have not presented data concerning the fingerprint examiners' accuracy. What little evidence concerning accuracy that does exist has generally been discussed in the literature by scientists and scholars discussing the absence of accuracy measurements (e.g., Cole, 2001, 2003, 2005a, 2005b, 2006, 2007; Haber & Haber, 2003; Kennedy, 2003; Loftus & Cole, 2004; Specter, 2002).

Before presenting our own research, we discuss three distinct investigations that attempted to assess how well different groups of FPEs could match photographs of latent prints to other photographs of prints of the source fingers under varying circumstances. The first investigation concerned what has come to be called a "proficiency test". According to some, a "proficiency test" is not a test of either validity or reliability. We will not debate the alleged distinction here, as it only strengthens the claim of no empirical evidence (but see Cole, 2006, for further discussion of the alleged distinction). For our purposes, this test was artificial: Although it involved "latent" prints and ten-print cards and used real FPEs or their labs, the task itself was an experiment, similar in that respect to the experiments we report here. The second investigation was equally artificial, but concerned the performance of FPEs at a functioning police forensics unit. The third investigation concerned the influence of contextual information on new judgements from FPEs to real-world, forensic fingerprint pairs previously assessed by the FPEs to be matches. None of these investigations comes close to being an adequate test of FPE abilities in general but, collectively, present little support for FPEs' claimed high level of accuracy (Ashbaugh, 1994) to match correctly two different prints from the same source finger. We present them here only as exemplars of the state of knowledge, not as true measures of FPE performance.

The CTS proficiency tests

The first case is provided by the CTS (Collaborative Testing Services, Inc.) "9508" test. Collaborative Testing Services, Inc. has conducted annual or semiannual proficiency tests of various forensic procedures, including fingerprint identification by FPEs, since 1983. CTS publishes the full (summaries and the raw data by FPE/lab) results of the more recent tests on its web site, but only the summaries of the results of the CTS fingerprint tests before 2001 are publicly available. We obtained the complete 9508 test results through personal contact with a scholar who had access to a paper copy of the published test.

The CTS 9508 test occurred in 1995, and data were collected from FPEs at 156 labs. Each lab was provided with a scenario that described a murder investigation. Seven "bloody" prints were left at the crime scene, and the FPEs were asked whether any of them could be identified to the "full-rolled", ten-print cards provided by four people: three potential suspects and the murder victim. Photographs of both the ten-print cards and bloody prints were provided. Of the seven prints purportedly left at the crime scene, five were "targets", three from one suspect (right index finger, left index finger, and right thumb) and two from the murder victim (both from the same finger—right ring). That is, for each of the five targets, the photograph of the "full-rolled" print from the same finger was on one of the four ten-print cards available as potential matches. The remaining two bloody prints were "distractors" or "lures" and, oddly, were provided by the twin (kind unspecified) brother of the second suspect, who was not a target and who provided the third ten-print card. The fourth (again, nontarget), ten-print card was apparently provided by a woman.

FPEs were allowed responses that ranged from identifying the corresponding print on a ten-print card, through declaring that the evidence was not sufficient to make a judgement, to declaring that no match existed on the four ten-print cards. For this and all subsequent reports, CTS compiled and summarized the data in various ways within the same report, sometimes as a function of the kind of source print for some statistics, sometimes as a function of the kind of correct response, and other times as a function of the kind of error, the proportion of FPEs/labs producing various responses, and so on. Unfortunately, no matter how one deconstructs these summaries, it is not possible to compute from them the relevant mean hit, misidentification, and false-positive rates.

The appropriate analysis is to produce mean rates per FPE—that is, to analyse the raw data with FPE (i.e., lab) as the unit of analysis, computing hit, misidentification, and false-positive rates over the targets and distractors for each FPE/lab, and then averaging. Accordingly, we rescored the original CTS 9508 data that way. Hits were defined as a correct match to the appropriate finger on the corresponding ten-print card of the five targets. False positives were defined as a claimed match to any finger on any of the ten-print cards of the two distractor prints. Misidentifications were defined as targets claimed to match an incorrect source finger. The results for the CTS 9508 task were a mean hit rate of 80.38%, a mean false-positive rate of 11.22%, and a mean misidentification rate of 7.69% (the standard errors of the mean were 2.19%, 1.79%, and 3.92%, respectively). This mean performance is not inconsistent with that of other matching tasks in the cognitive and applied psychological literatures with (different) photographs of the same or highly similar objects (e.g., unfamiliar faces, Megreya & Burton, 2006) more generally.

The CTS 9508 is probably the best example of a FPE proficiency test in the literature; however, it is still quite flawed. The performance results are critically dependent on the exact print exemplars used both as targets and, equally, as distractors. With such a limited set of latents, even one highly distinctive distractor, for example, could significantly improve discrimination by reducing false positives, whereas a single highly distinctive target could vastly improve performance by increasing hits. Furthermore, this problem is exacerbated if every

FPE/lab views the same targets as targets and distractors as distractors. At a minimum, the latents used should be counterbalanced over FPEs/labs as to being a target or a distractor. Better, would be to have different samples of targets and distractors, counterbalanced over FPEs/labs, that differ in discriminability in known ways.

Performance on subsequent CTS proficiency tests has been in some sense much better (often with hit rates well over 90%, and with false-positive rates often in the low single digits or even less, where calculable), but again, with very few (and fixed) latents of unknown distinctiveness. As there is no reason to assume that the selection of labs and FPEs had been biased toward those who were more competent or diligent than those in the 9508 test, perhaps the best we can conclude about these subsequent CTS tests is that they were certainly "easier" in some sense for the FPEs/labs than the 9508 test. Still, in the absence of (preferably) large, counterbalanced samples of latents as targets and distractors, and/or knowledge of their individual distinctiveness, it is not clear what these tests say about FPE proficiency in fingerprint identification, if they say anything at all.

Boston Police Department Latent Print Unit tests

A recent investigation of the members of the Boston Police Department Latent Print Unit by Ron Smith and Associates in 2004 provides a convenient example of the problem.[1] Two proficiency tests were administered, each involving a different set of 15 latent prints. Outside experts evaluated all of the latent prints as being sufficient for identification or exclusion. However, the 15 latents in the second set were assessed to be "of more advanced levels of difficulty" than those in the first set, although the bases of this assessment were not reported. In the test with the first (easy) set, all 15 prints were "targets"; as with many such tests, no distractors were used. Five members of the Unit participated. Our analyses found a mean hit rate of only 50.67% (and a standard error of the mean of 7.77%).

Four of the five members also participated in the second ("difficult") test. In this case, one of the 15 latents was a distractor. Our analyses of these data found that the mean hit rate in this case was only 19.0% (standard error of the mean was 9.1%), with a false-positive rate of 25% (standard error of the mean was 25%). That is, the Unit members were unable to discriminate the 14 targets from the 1 distractor for the "difficult" test, despite the outside expert claim that the detail available in the latent prints was sufficient for identification or exclusion. Clearly, as with the CTS tests, some latent prints in some contexts are more difficult to discriminate than others despite there being claimed sufficient detail in the latent for making just such judgements.

Contextual influences on FPEs

Dror, Charlton, and Péron (2006) investigated whether the context under which a judgement was to be made would influence FPEs' decisions. Five examiners were each asked to assess whether a pair of prints was a match (i.e., from the same source finger) or not. Each pair was unique to each FPE and was, in fact, undeclared to the individual examiner, a pair that the examiner had previously determined to be a match in the normal course of an actual criminal investigation. Two independent FPEs confirmed that the five pairs were indeed all matches. When the five FPEs were asked to examine these same print pairs in the context of each of them being from a notorious nonmatch case (the FBI's erroneous identification of the Madrid bomber; Fine, 2006), only one of the five still declared the print pair to be a match; one now concluded that it was not possible to decide, and the remaining three concluded that the pairs were now nonmatches.

In a related experiment, Dror and Charlton (2006) investigated similar sources of error in fingerprint identification with another six highly

[1] A copy of this report may be obtained by writing to the author SAC.

experienced FPEs. Unbeknownst to each FPE, each received eight sets of prints from the expert's own prior judgement history, one half in a biasing context contrary to the expert's prior judgement and the remaining sets in an unbiased context. Orthogonal to that distinction, one half of the sets were prior identifications (i.e., claimed hits) and the remainder claimed exclusions. Finally, orthogonal to each of the prior distinctions, one half of the items were judged by two independent experts as relatively more difficult assessments and the remainder as relatively simpler. As in the previous experiment, the FPEs here were found to produce assessments inconsistent with their prior assessments, even on the unbiased fingerprint sets, although not at rates as high as those in the previous experiment.

These few investigations with their very limited (and usually unspecified) latent print distributions probably tell us more about the specifics of the prints and testing format used in these tests than about the proficiencies of the fingerprint examiners. Still, on the evidence, especially the just-cited work of Dror et al. (2006) and Dror and Charlton (2006), it is clear that, at least under some circumstances—even when the prints are assessed by other FPEs to be within the expertise of FPEs—they did not necessarily perform exceptionally well. The caveat is important: All of the tests just described were thought by at least a subset of FPEs to be of the kind that fingerprint experts (i.e., those trained, and often court certified, to make such judgements) should be capable of assessing reliably. Still, there is no reason to believe that these few, extremely limited investigations provide much evidence regarding the skills and performances of FPEs more generally.

To assess FPEs' or anybody's ability to identify fingerprints, it is necessary to perform an experiment. Any such experiment will necessarily raise questions about how faithfully it replicates the expert's task as performed in actual casework. In the case of fingerprint identification, forensic analyses appear to fall into two basic types. In some cases, a latent print examiner is presented with a latent print or prints and a closed set of known prints from a suspect or suspects (suspect-present analysis). The analyst will then determine whether the latent print shares a common source with one of the known prints. In other cases, however, the FPE may be presented with a latent print or prints and no known prints from suspects (suspect-absent analysis). In both types of cases, the analyst may also be presented with known, "elimination" prints from victims, "bystanders" with legitimate access to the scenes, or police personnel who attended the scene. Here, the analyst will typically perform a computer-aided database search that will generate a list of candidate matches, prints from persons in the database that bear friction ridge skin arrangements most similar (as given by the computer algorithm) to that found on the latent print. The analyst will then proceed through the candidate list to determine whether any of them shares a common source with the latent print. Although we are aware of no data that specify FPEs' division of labour between these two types of cases, it is clear that historically FPEs mostly worked cases of the former type; database searches before the dissemination of computerized database searching aids were reserved for rare, high-priority cases. Due to advances in computing, however, it is to be expected that the proportion of cases that fall into the latter type has been rising and will be expected to continue to rise (Cole & Lynch, 2006).

Each type of case presents its own opportunities and challenges. In suspect-present cases, the examiner essentially knows that the known prints derive from individuals who are suspected of being the source of the latent print for other reasons. Therefore, the analyst may be vulnerable to context bias (Dror & Charlton, 2006; Dror et al., 2006). On the other hand, a randomly selected distractor has a low likelihood of being very similar to the true source just by chance, which would seem to reduce the likelihood of false positives. In suspect-absent cases, the opposite situation obtains: The computer searches the database for those known prints most similar to the latent print. If the true source is present in the database, it is highly likely (though not certain) that it will appear on the candidate list. At the same time, however, the computer is selecting the distractors most like the true source and presenting them to the analyst. In other words, the database search is maximizing the difficulty of FPE's task and maximizing the conditions conducive to false positives.

These two types of search involve quite different distractors. In suspect-present cases, distractors are chosen essentially at random. It is unlikely that they will strongly resemble the true source just by chance, but, if they do, they may be subject to context bias. In computer-aided, suspect-absent cases, the search provides the most difficult possible distractors (at least in so far as the search algorithm is concerned). Because FPEs' duties involve both tasks and because no data are available concerning the relative distribution of these tasks, it is not immediately obvious to the researcher how distractors should be selected for experimental versions of these tasks. As we demonstrate, the choice of distractors does have an effect on accuracy.

From a psychological perspective, however, errors in such visual judgements are to be expected (e.g., Dror & Charlton, 2006; Dror et al., 2006). Given the all-too-human foibles of distraction, lapses of attention, fatigue, rushes to judgement, less than perfect information, biases, expectations, and so on that cannot be avoided even by the most diligent, errors will happen. Given the inevitability of errors, then, just how difficult is the task of matching fingerprints, and what are the factors that influence that difficulty?

The current research is a preliminary attempt to answer such questions using naïve undergraduates; indeed, the question would seem to be the perfect target of a psychophysical investigation. As noted, as with many, preliminary, psychophysical experiments, the current research is limited in its real-world generalizability. It ignores, for example, all the important questions about whether and how well various techniques can usefully lift latent prints from various surfaces, and it accepts for the sake of moving forward the supposition that the ridge pattern of every finger of every person who has existed or currently exists is unique. It also, and pointedly, ignores the possibility of training and expertise in fingerprint identification. Instead, it works the problem from the other end: Given that an image of a fingerprint has been obtained (however flawed, partial, and otherwise distorted it may be), how well can it be matched by common visual inspection with another image of a fingerprint (obtained the same way, in this case) from the same finger?

EXPERIMENT 1

This first experiment was an attempt to simulate a forensic context in which images of 20 prints—the crime scene "latents"—were to be compared with the ten-print cards of 10 elimination individuals and 2 "suspects". Unlike the "proficiency tests" discussed earlier, however, each participant received a unique, randomly selected set of target and distractor stimuli, distributed equally over each of the five digits of each hand.

Method

Participants

A total of 19 male and 29 female psychology undergraduate students and members of the broader University of Lethbridge community participated in the experiment.

Materials

Fingerprints were obtained from 127 members of the broader university communities of the University of Lethbridge and McMaster University. Each individual was asked to produce a print of each of 10 fingers in appropriately labelled boxes on a sheet of white paper. They were instructed to "ink" each finger pad by pressing it to a semi-inkless, professional fingerprint compound and then pressing the finger to the appropriate box on the paper. So that the obtained prints would be more like the "latent" prints obtained forensically, no further instructions were provided; in particular, individuals were not instructed to provide full "rolled" prints. Thus, some individuals produced relatively complete prints, others just partials; some pressed hard, producing dark and sometimes smudged prints, others light, ill-defined partials, and so on. Furthermore, some individuals attempted to provide prints in a canonical (upright) orientation to the box, others angled the prints, typically varying the angle with the finger printed, and so on. After providing a complete set of prints, each individual was asked with no forewarning to do so again on a new sheet.

Thus, from each individual, 20 fingerprints were taken, 2 from each finger, obtained in ten-print/finger sets. The prints were scanned as 8-bit, grey-scale computer graphic image files at 300 ppi, with each print centred and cropped to a standard rectangular image of 300×200 (rows $\times$ columns) pixels. The first set of images from each individual provided the source of the "unknown latents" to be matched, whereas the second set provided the source of the complete "test sets" from which matches to the unknown "latents" were to be selected. Note, however, that all of the print images are *patent* prints, not "latent" prints, as no special process was required to reveal them. Note also that in most forensic contexts, the test sets or "ten-print cards" are of "full-rolled" print images (often including palm prints and ridge details from other finger joints than just the finger tip), not just other patent partials of finger tips as used here. There were 1,270 images—127 complete person sets—in each set.

Procedure

For each participant in the experiment, 20 person sets were sampled at random from the unknown latents. One print was selected at random from each chosen person set subject to the constraints that there would be two prints from each of the 10 fingers: one to serve as a target, and the other as a distractor. The corresponding test sets of the 10 target prints, plus a further two sets selected at random from those remaining among the test sets provided the 12 full test sets.

Participants were given a cover story about a break-in at a professor's office from which 20 latent prints (the "unknown latents") had been obtained. The story explained that complete sets of prints (the "test sets") had been obtained from two suspects and 10 elimination individuals (e.g., the professor, her graduate students, research assistants, and colleagues) and that the participant's task was to attempt to find matches to each of the 20 unknown latents from within the 12 test sets. As would (or should) be true in the real world of forensic latent fingerprint identification, participants were told that the test sets were coded (so that "suspect" sets were not differentiated from "elimination" sets to the participants) and that matches were unlikely.

Each unknown latent print was presented individually as an image at approximately four times normal size on the left-side centre of the computer screen, accompanied by its coded label (e.g., "Latent 1"). A similarly scaled image of a test print appeared on the immediate right-side centre of the screen, abutting the latent image, accompanied by a label indicating the number of the set (1 through 12), hand, and finger it was from. This test print image could be rotated left or right (by clicking appropriately labelled buttons immediately below it) about its centre in single degree steps as many degrees as the participant desired. A 1-bit (black/transparent) "overlay" of the test print image at the same orientation could be obtained by clicking on the test image. This overlay could then be dragged over the latent and could be further rotated by continuing to rotate the source image. Across the bottom of the screen were images at approximately normal size of the 10 prints from the current test set, labelled accordingly. Each of these images could be clicked on to have it appear as the test image. Doing so would not only replace the current test image, but would reset the orientation to the canonical angle and remove any overlay that may have been generated from the previous test print. Below these images were buttons to move backward or forward among the test sets, resetting the images and labels in the process.

On the extreme right of the screen were two further buttons, labelled "Match" and "No Match". Clicking on the "Match" button presented a record on the left side of the screen of the current choice (the set, hand, and finger of the current test image, and that a match had been declared), with a pop-up button labelled "confidence" below it. Similarly, clicking on the "No Match" button presented a record that a nonmatch had been declared, with again, a confidence button below it. Clicking on the confidence button allowed participants to select a confidence from 1 (low confidence) through 6 (high confidence) in their match or nonmatch decision and presented a further button labelled "next latent". Participants were free to continue to look for and select other possible matches for the current test print or to decide a nonmatch, changing the displayed information and requiring a new confidence judgement in the process. Clicking the "next latent" button would record the details of the last choice made and present the next unknown latent for judgement.

Results and discussion

Match responses on target-present trials give rise to correct identifications, or hits (set, hand, and finger correct), and misidentifications (some other set, hand, or finger misidentified as the target). On target-absent (i.e., distractor) trials, they provide false-positive responses. The mean proportions of each of these three match responses as a function of source finger type (i.e., thumb through little finger, collapsed over hand) are shown in Figure 1.

As may be seen in Figure 1, prints from index fingers had a higher mean hit rate and lower mean misidentification and false positive rates than those from the other source finger types. Hit and misidentification rates are not mathematically independent as they are obtained as contrasting responses to the same targets and, hence, cannot meaningfully be analysed together as a

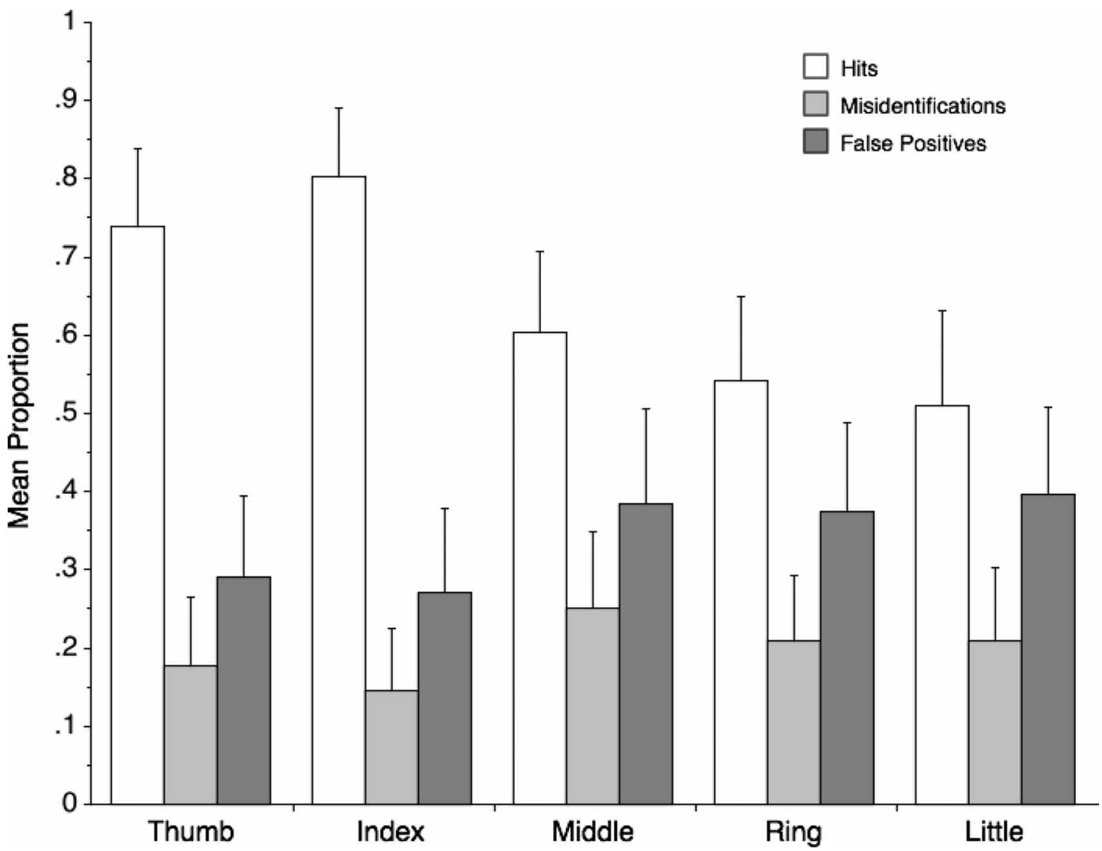

Figure 1. *Mean proportion of hits, misidentifications, and false positives as a function of source finger type of the print in Experiment 1. Error-bars are within-cell 95% confidence intervals about the mean.*

within-participant factor. However, as they do represent a serious forensic error, the mean misidentification rates were analysed separately as a function of source finger type in a one-way, within-participant analysis of variance (ANOVA), with participants crossing source finger type as the random variate. An $\alpha = .05$ was used as the level of significance for all comparisons in this paper. The overall mean rate (i.e., proportion) of misidentifications of .20 did not vary significantly as a function of source finger type, $F(4, 188) = 1.40$, $MSE = 0.05$.

The remaining "match" responses to targets (hits) and distractors (false positives) were analysed as a function of source finger type. A 2 (hits vs. false positives) $\times$ 5 (source finger type) within-participants ANOVA was conducted with participants crossing both factors as the random variate. The mean hit rate of .64 significantly exceeded the mean false positive rate of .34, $F(1, 47) = 21.94$, $MSE = 0.48$. However, although there was no significant variation in claimed matches as a function of source finger type, $F(4, 188) = 1.30$, $MSE = 0.10$, source finger type did interact significantly with the discrimination of targets from distractors, $F(4, 188) = 10.58$, $MSE = 0.06$, confirming the impression from Figure 1 that some source finger types are easier to discriminate via matching than are others. Analysing hits and false positives independently revealed that the hit rate for thumbs was significantly higher than that of middle, ring, and little fingers (Fisher's $LSD_{.05} = 0.123$ for all comparisons). The hit rate for index fingers was also significantly greater than that for middle, ring, and little fingers. No other hit rate comparisons were significant. The false-positive rate for index fingers was significantly lower than that of middle and little fingers (Fisher's $LSD_{.05} = 0.112$ for all comparisons). No other false-positive rate comparisons were significant. Thus, thumbs and especially index fingers were better discriminated than any of the remaining source finger types, which did not differ among themselves. That the finger type (e.g., index vs. little finger) affects the identification of the source finger of prints provides clear evidence that some prints are easier to identify than others.

EXPERIMENT 2

Experiment 1 attempted to simulate a real-world, suspect-present forensic context, allowing for the full range of match responses—including misidentifications—but thereby leaving the resulting analysis somewhat problematic. Furthermore, the task was complicated by the common, real-world requirement that the participant search through many panels of 10-choice images of patent prints for the one matching (or not) image of a print from the same source finger. In Experiment 2, the task was reduced to the last stage of the matching process, a simple yes/no or match/no-match task. Despite the superficial simplicity, the task is both forensically relevant and, indeed, a common forensic task as noted in the introduction. Oftentimes, FPEs are asked whether two latents originated from the same source finger, or, following a winnowing of possible matches from an elimination search of a database of fingerprint images, whether a latent can be matched to a source image from the database.

In the current experiment, on each trial, a participant was presented with just two images of prints and was asked to judge simply whether the two prints were from the same source finger or not. Hence, in this case, a hit was a declaration of a match when the two print images derived from the same source finger, and a false positive was the incorrect declaration of a match when the two print images differed as to source finger. Misidentifications were not possible, as hits required only a "match" response, rather than the further selection of the correct print image as was the case for Experiment 1.

Method

Participants

A total of 4 male and 24 female psychology undergraduate students and members of the broader University of Lethbridge community participated in the experiment.

Materials
The materials were the same 127 fingerprint sets as those used in Experiment 1.

Procedure
For each participant, 80 fingerprint sets were sampled at random from the 127 sets. One half of these sets were selected at random to serve as the target or "match" trials, and the remainder as the distractor or "nonmatch" trials. For the distractor trials, a further 40 sets (one for each distractor trial) were selected at random from those remaining to provide the nonmatch source fingers. One distractor image was provided by the source finger from the designated set, and the other was provided by the corresponding finger from the corresponding nonmatch set (i.e., the same finger from a different person). From the 40 target and 40 distractor sets, 4 sets were selected at random to provide trials for each of the 10 (right vs. left hand by five fingers) source finger types. Thus, for each of 10 source finger types, there were 4 match or target trials and 4 nonmatch or distractor trials, for a total of 80 test trials.

The 80 trials were presented in a random order for each participant. For each trial, the source finger from the designated set was presented on the left of the display at roughly four times normal size, as in Experiment 1. The match or no-match image was displayed on the right. As in Experiment 1, participants could rotate the image on the right of the display and create and rotate a black/transparent overlay that could be dragged over the image on the left of the display. Below the fingerprint images was a 12-point scale, labelled from "Sure Different", "Guess Different", through the midpoint, to "Guess Same" and "Sure Same". On the scale was a slider that the participant used to indicate the same/different judgement and confidence for each trial. Participants were free to adjust the slider until they clicked a button below the scale labelled "Done", which then saved the decision and confidence for the current trial, and advanced to the next trial.

Results and discussion

Hits are declarations of a match to match or target trials, and false positives are declarations of a match to nonmatch or distractor trials. The mean hit and false-positive rates as a function of source finger type are shown in Figure 2. These data were subjected to a 2 (hits vs. false positives) × 5 (source finger type collapsed over hand) within-participant ANOVA with participants crossing both factors as the random variate. Participants significantly discriminated match trials from nonmatch trials, $F(1, 27) = 7.92$, $MSE = 0.05$, and at hit and false-positive rates similar to that found with similar visual matching tasks with unfamiliar stimuli (e.g., unfamiliar faces, Megreya & Burton, 2007). No source finger type attracted significantly more positive responses than any other, $F(4, 108) < 1$, but, as may be seen in Figure 2, some source finger types were better discriminated than others, $F(4, 108) = 8.77$, $MSE = 0.01$, replicating the similar result in Experiment 1.

The hit rates of the thumb, index, middle, and ring fingers were each significantly greater than that of the little fingers (Fisher's $LSD_{.05} = 0.057$ for all comparisons) and did not differ significantly from one another. The false-positive rates for index, middle, and ring fingers were also significantly lower than that of the little fingers (Fisher's $LSD_{.05} = 0.062$ for all comparisons), but that for thumbs was not. Furthermore, the false-positive rate for thumbs was significantly higher than that for index fingers. No other comparisons were significant.

These differences in discrimination as a function of source finger type are perhaps better appreciated by signal detection theoretic analyses of the data. In Figure 3, depicted are the receiver-operating characteristic (ROC) curves fitted to the mean confidence-scale responses to each of the five source finger types in Experiment 2. ROC curves plot the unit square of paired hit and false-positive rates for different criterion settings of the willingness to emit "yes" or, in this case, match responses. They were derived for each participant (and then averaged) based on

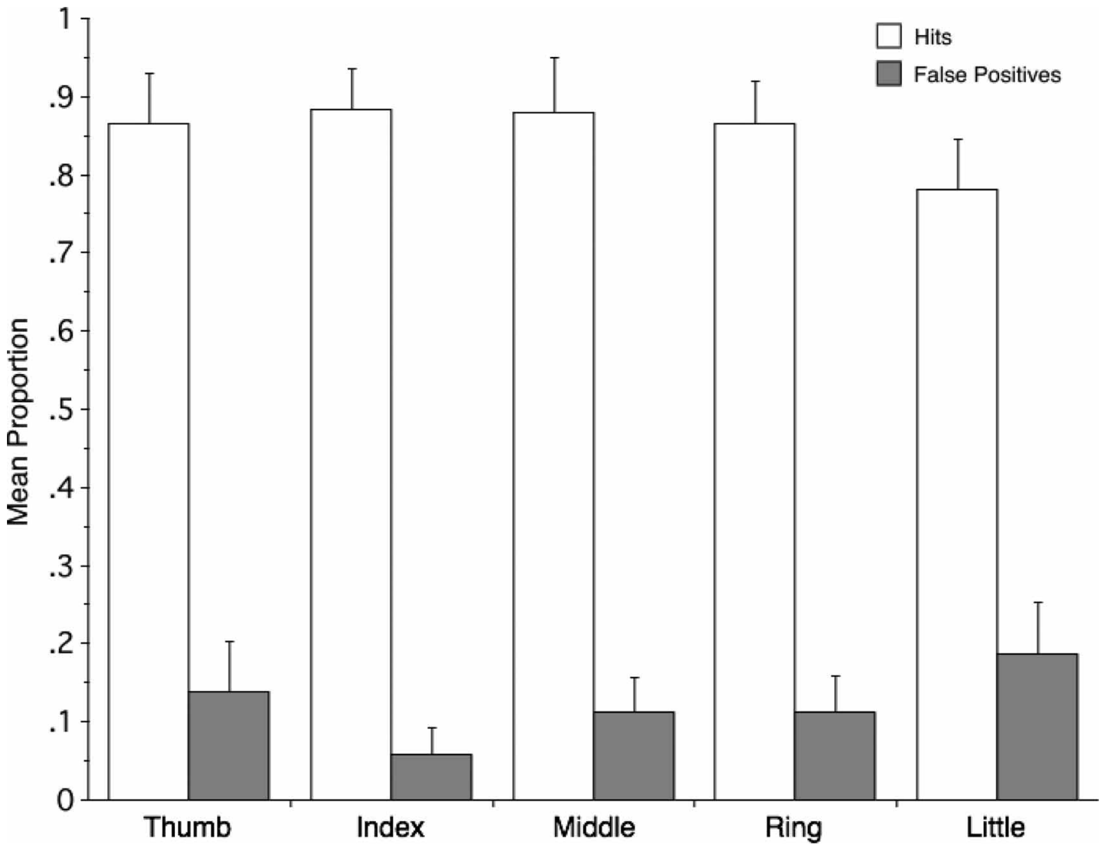

Figure 2. *Mean proportion of hits and false positives as a function of source finger type of the print in Experiment 2. Error-bars are within-cell 95% confidence intervals about the mean.*

the confidence levels assigned to each response (see, e.g., Wickens, 2001). The fitted curves were computed via a web-based program (see Eng, n.d.). As can be seen, test pairs of little fingers are clearly discriminated less well than are those from other source finger types, with, again, index fingers evincing the best discrimination. Pad size of the print apparently is not the sole reason for superior discrimination, as, among the remaining source finger types, thumbs (the largest finger-pad for most people) were less well discriminated than were index finger types. This general pattern was confirmed by an analysis of the discrimination index, A', derived from participants hit and false-positive values. Shown in Table 1 are the mean A' values for Experiment 2 as a function of source finger type. These data were subjected to a one-way (source finger type) within-participant ANOVA, with source finger type crossing participant as the random variate. There was a significant effect of source finger type, $F(4,\ 108) = 5.34$, $MSE = 0.005$; thumbs, index, ring, and middle fingers were discriminated

Table 1. *Mean A' values for Experiments 2 and 3*

		Source finger				
Experiment	*Similarity*	*Thumb*	*Index*	*Middle*	*Ring*	*Little*
2	Random	.91	.95	.94	.93	.87
3	Low	.83	.93	.91	.94	.8
	High	.93	.96	.91	.92	.8

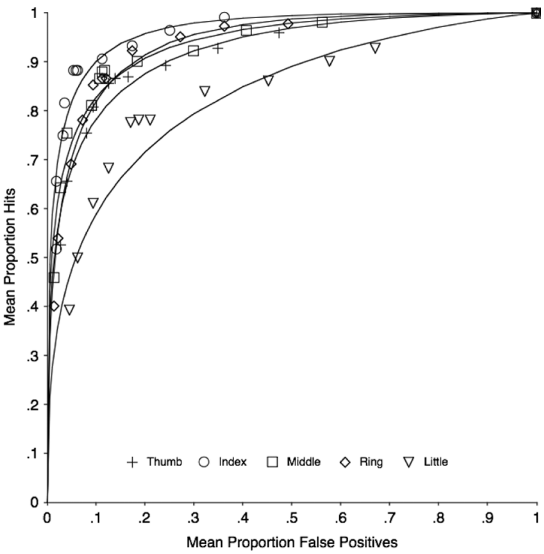

Figure 3. *Mean receiver-operating characteristic (ROC) hit and false-positive values derived from confidence judgements and the corresponding, fitted ROC curves (assuming equal-variance, Gaussian distributions) as a function of source finger type of the fingerprints in Experiment 2.*

significantly better than were little fingers (Fisher's $LSD_{.05} = 0.037$ for all comparisons), and thumbs were less well discriminated than index fingers. No other comparisons were significant.

EXPERIMENT 3

In Experiments 1 and 2, each participant was exposed to a random selection of target and distractor prints equally from all five digits of each hand, allowing for an assessment of discrimination as a function of source finger type. Performance varied considerably with the particular digit examined, sometimes on hits and other times on false positives. For example, the superior hit rate found for index fingers relative to some of the others may be due to index fingers providing for prints that are generally more alike than for those from, say, thumbs. Similarly, the higher false-positive rate for little fingers may be due to their providing for pairs of distractor prints that are also generally more alike than those from index fingers. There is no reason to assume that the sources of similarity are necessarily the same. For example, because index fingers are used more frequently than are the others, they may be more likely to acquire distinguishing characteristics, possibly making them easier to match, and to distinguish as distractors. Whereas, little fingers may be more difficult to distinguish from one another simply because their small pad size provides for little distinguishing detail.

As noted in the introduction, not all (and recently perhaps not even most) fingerprint forensic tasks involve anything like a random distribution of distractors. Indeed, a computer-based search provides a set of all prints from the forensic database that are sufficiently similar (in terms of the computer algorithm) to a target to be considered fungible candidates. As at most only one of these candidates could be a match to the target, the remaining candidates provide a set of highly similar distractors that is anything but random. Such a distribution of distractors may have serious consequences for FPE performance possibly by significantly elevating the rate of false positives. It also may have consequences for hit rates, as the expertise of forensic examiners will be increasingly called upon to assess only the more difficult of cases, such as those concerning partial, degraded, or low-quality latent prints where even correct matches may be problematic to discern for computer-based algorithms.

Neither of the previous experiments investigated the effect of such difficult target and distractor sets. To conduct such an experiment requires some metric for assessing the similarity among images of prints. FPEs are trained to assess similarity by isolating the minutia or Galton "points" of otherwise globally similar prints (e.g., whorl or arch patterns) that are held in common by location (Cole, 2001). The more naïve participants in our experiments, on the other hand, may do something similar, or may assess similarity in a wholly different way that may or may not be correlated with the assessments of the FPEs. Accordingly, any one metric is unlikely to capture all of the ways in which one print can be regarded as similar to another.

However, an approach that has proved itself useful for visual stimuli of many kinds is to define similarity as the proximity of one stimulus to another in a highly multidimensional image space. One technique that uses this approach is the principal component analysis (PCA) of pixel-maps. Although this technique has its home in photographs of faces (e.g., Abdi, Valentin, Edelman, & O'Toole, 1995; Burton, Bruce, & Hancock, 1999; O'Toole, Abdi, Deffenbacher, & Valentin, 1993; Turk & Pentland, 1991), it has been used successfully with several other kinds of stimuli (Vokey, 2001) such as the structure of photographs of natural scenes (Baddeley & Hancock, 1991; Heidemann, 2006), style of artist paintings (Vokey & Tangen, 2001), kin recognition of photos of chimpanzee faces (Vokey, Rendall, Tangen, Parr, & de Wall, 2004), and even such esoterica as the structure of strings from artificial grammars (Vokey & Higham, 2004). In each case, stimuli are projected into the multidimensional space of all such stimuli to return a vector, and the similarity of one stimulus to another is given by the cosine of the angle between their vectors: Cosine values close to 1.0 indicate that, at least as defined by that space, the stimuli are virtually identical; whereas cosines close to zero indicate that the stimuli are highly dissimilar. Because the analysis is at such a low-level description of the stimuli (pixel-maps), it is likely to capture similarity arising from either micro or holistic features. Details of the method are available from any of the just-cited works.

Method

Participants

A total of 22 male and 32 female psychology undergraduate students from the University of Lethbridge participated in the experiment.

Materials

The materials were the same as those used in Experiments 1 and 2. The first members of each pair of prints from the same source were used to construct the multidimensional PCA space. The second print from each pair was then projected into that space. For the first member of each pair of prints from the same source finger, the cosine similarity between it and all prints from the second members of each pair of prints from the same source finger type was then computed. Thus, we obtained the cosine similarity of each print to its correct match and also both the highest and lowest similarity nonmatch print from the same source finger type. So, for any given target print, there were three prints with which it could be paired at test: its correct match and, from the remaining 126 prints from the same source finger type, its lowest similarity nonmatch print and its highest similarity nonmatch print. Target prints were rank-ordered by their cosine similarity to their correct match and, hence, could be designated as either high- or low-similarity matches based on the end of that spectrum from which they were selected. Thus, we have available for the experiment both between-print high-similarity and between-print low-similarity matches and, within any given target print, both high- and low-similarity nonmatches. A total of 80 target prints were selected to be used in the experiment: the first 40 from the high-similarity matches and the last 40 from the low-similarity matches.

Procedure

Across all participants, materials were counterbalanced such that a given high- or low-similarity match print was used to provide either a match trial or a nonmatch trial, and, if used to provide a nonmatch trial, it was paired with either a high- or a low-similarity nonmatch print as described from the same source finger type. As in Experiment 2, each participant received 80 trials in random order, one half of which were matched trials (targets) and one half of which were nonmatch trials (distractors). All other aspects of the procedure were also the same as those in Experiment 2.

Results and discussion

The mean hit and false-positive rates as a function of source finger type and similarity of the test pair (low vs. high) are shown in Figure 4. These data were subjected to a 2 (hits vs. false positives) × 2

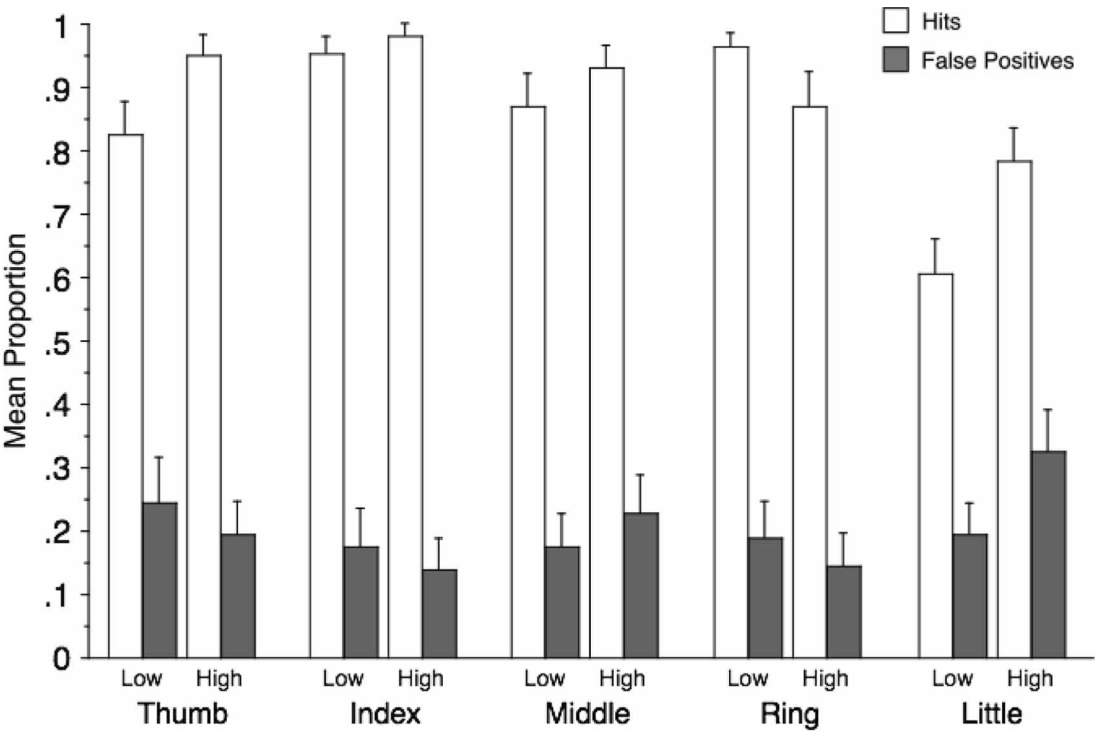

Figure 4. *Mean proportion of hits and false positives as a function of source finger type of the print and the similarity of the test pair in Experiment 3. Error-bars are within-cell 95% confidence intervals about the mean.*

(similarity of the test pair: low vs. high) × 5 (source finger type) within-participant ANOVA with participants crossing all three factors as the random variate. Participants significantly discriminated match trials from nonmatch trials, $F(1, 53) = 1{,}419.90$, $MSE = 0.09$. Some source finger types attracted significantly more "match" responses than did others, $F(4, 212) = 11.18$, $MSE = 0.02$, and, as in the previous experiments, some source finger types were better discriminated than others, $F(4, 212) = 34.42$, $MSE = 0.03$.

Of greater interest for the current experiment are the effects of the similarity of the test pairs. There was a main effect of similarity: As would be expected, low-similarity pairs attracted significantly fewer "match" responses than did high-similarity pairs, $F(1, 53) = 7.12$, $MSE = 0.04$, but this effect interacted significantly with whether the trial was a match or nonmatch trial, $F(1, 53) = 8.27$, $MSE = 0.02$. Simple effects analyses on this interaction revealed that low-similarity pairs reduced hits significantly relative to high-similarity pairs, $F(1, 53) = 23.24$, $MSE = 0.02$, but had no effect on false positives, $F(1, 53) < 1$. One might have anticipated a mirror effect pattern here: a positive effect of similarity on hits and a negative effect on false positives. However, for unfamiliar stimuli, such as unfamiliar faces, in paired matching tasks, Megreya and Burton (2007) have demonstrated that there is no necessary relation between hits and false positives: Factors that affect hits need not have any corresponding influence on false positives. Similarity also interacted significantly with source finger type, $F(4, 212) = 17.01$, $MSE = 0.02$, indicating that the effect of the similarity of the test pair varied for different source finger types in terms of their ability to attract "match" responses. Simple effects analysis of this interaction revealed that there was no significant effect of similarity for both thumb, $F(1, 212) = 3.49$, $MSE = 0.02$, and index, $F(1, 212) < 1$, finger test pairs, but both middle, $F(1, 212) = 7.84$, $MSE = 0.02$, and little, $F(1, 212) = 59.32$, $MSE = 0.02$, finger test pairs attracted significantly more "match" responses for high-similarity than low-similarity test pairs, as might be expected for effects of similarity. However, ring finger test pairs attracted significantly more "match" responses for low- rather than high-similarity test pairs, $F(1, 212) = 12.26$, $MSE = 0.02$; we have no idea why, but it does confirm that interprint similarity plays an important, and possibly complicated, role in fingerprint identification. As can be seen in Figure 4, the degree of complication is borne out by the fact that this effect itself interacted with whether the trial was a match or no-match trial, $F(4, 212) = 13.57$, $MSE = 0.03$. Signal detection analyses were used to deconstruct these complicated effects.

Shown in Figure 5 are the mean ROC curves for the discrimination of match from nonmatch test pairs computed as in Experiment 2, for both the low-similarity (Figure 5A) and high-similarity (Figure 5B) test pairs. In general, high-similarity test pairs were better discriminated than were low-similarity test pairs. However, as in Experiment 2, little finger test pairs were clearly less well discriminated than were the prints from other source finger types, and there is some suggestion that low-similarity, thumb test pairs, in particular, were less well discriminated than were index, middle, and ring finger test pairs. To confirm these trends, an analysis of the discrimination index, A', derived from participants' hit and false-positive rates, was performed. Shown in Table 1 are the mean A' values as a function of source finger type and similarity of the test pairs for Experiment 3. These data were subjected to a 2 (similarity of the test pair: low vs. high) × 5 (source finger type) within-participant ANOVA with participants crossing both factors as the random variate. As already noted, test pairs from some source finger types were better discriminated than were others, $F(4, 212) = 26.66$, $MSE = 0.01$. High-similarity test pairs were significantly better discriminated than were low-similarity test pairs,

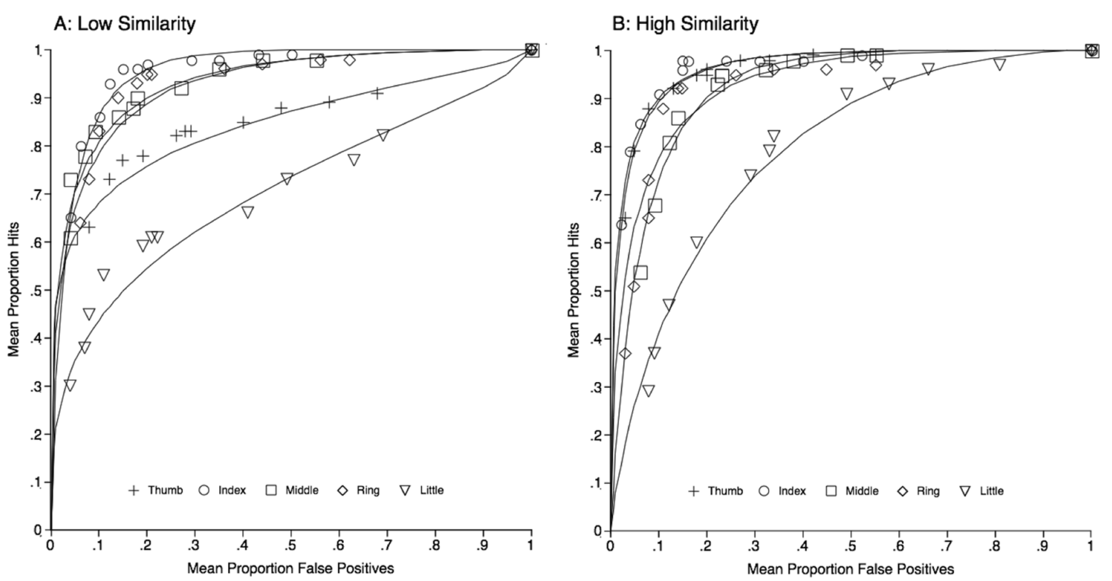

Figure 5. *Mean receiver-operating characteristic (ROC) hit and false-positive values derived from confidence judgements and the corresponding, fitted ROC curves (assuming equal-variance, Gaussian distributions) as a function of the similarity of the test pairs and source finger type of the fingerprints in Experiment 3.*

$F(1, 53) = 8.75$, $MSE = 0.01$, but this effect was compromised by a significant interaction with the source finger type, $F(4, 212) = 4.59$, $MSE = 0.01$. As is clear from Table 1, simple effects analyses revealed that only the discrimination of thumb test pairs was differentially affected by test pair similarity, $F(1, 212) = 21.47$, $MSE = 0.01$, suggesting that the whole of the main effect of test pair similarity was carried by this effect on the discrimination of thumbs.

GENERAL DISCUSSION

In the introduction we questioned the difficulty of fingerprint identification and asked what factors might influence that difficulty. The results of our experiments support three major conclusions in answer to these questions. First, people can clearly identify fingerprints quite well. Second, this ability varies as a function of the source finger type (e.g., index vs. little fingers): The prints from some fingers are more difficult to discriminate than are others. Third, it also varies as a function of the interprint similarity of pairs to be identified, particularly for thumb prints.

In the first experiment, participants were presented with a series of unknown "latent" fingerprints that were purportedly left at a crime scene and a collection of prints from suspects and elimination individuals. Their task was to match the prints from the crime scene to their source, thereby simulating an authentic forensic scenario. Because each participant was exposed equally to a random selection of prints from all five digits of each hand, we could assess discrimination as a function of source finger type. Performance varied considerably with the particular digit examined. Thumb and index impressions were identified more readily than those from little fingers.

Experiment 2 was directed at the final stage of the identification process: a basic matching task, which, in addition, provided for a signal detection analysis of participants' performance. Using the same random assignment of prints from each finger of each hand as in Experiment 1, participants were simply asked to indicate how confident they were that a pair of prints matched or not. Discrimination performance varied with source finger type.

Experiment 3 was designed as a replication of Experiment 2, except that the interprint similarity of both match and nonmatch trials was varied. The results confirmed that discrimination performance varied with source finger type (e.g., in all three experiments, little fingers have been more difficult to discriminate). They also demonstrated that interprint similarity affects participants' ability to identify prints, especially thumb prints.

Whether these conclusions extend to the identification performance of FPEs is of course unknown. To do so requires that the requisite experiments with FPEs be performed, and, as noted in the introduction, in the 100 years of the forensic use of fingerprint identification, no such experiments have been reported. But even in the absence of such welcome experiments, it is clear from these laboratory results that people generally have substantial abilities to identify fingerprints despite what some of the studies reported in the introduction might be taken to imply. In that regard, our results suggest that subsequent field work with FPEs may wish to target such factors as the source finger type and explicit measures of interprint similarity.

Original manuscript received 6 August 2007
Accepted revision received 20 June 2008
First published online 25 November 2008

REFERENCES

Abdi, H., Valentin, D., Edelman, B., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception*, *24*, 539–562.

Ashbaugh, D. R. (1994). The premise of friction ridge identification, clarity, and the identification process. *Journal of Forensic Identification*, *44*, 499–516.

Baddeley, R. J., & Hancock, P. J. B. (1991). A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings of the Royal Society B*, *246*, 219–223.

Burton, A. M., Bruce, V., & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, *23*, 1–31.

Cappelli, R., Maio, D., Maltoni, D., Wayman, J. L., & Jain, A. K. (2006). Performance evaluation of fingerprint verification systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 3–17.

Cherry, M., & Imwinkelried, E. J. (2006). A cautionary note about fingerprint analysis and reliance on digital technology. *Judicature*, *89*, 334–338.

Cole, S. A. (1998). Witnessing identification: Latent fingerprint evidence and expert knowledge. *Social Studies of Science*, *28*, 687–700.

Cole, S. A. (2001). *Suspect identities: A history of fingerprinting and criminal identification*. Harvard, MA: Harvard University Press.

Cole, S. A. (2003). Fingerprinting: The first junk science? *Oklahoma City University Law Review*, *28*, 73–92.

Cole, S. A. (2004). Grandfathering evidence: Fingerprint admissibility ruling from *Jennings* to *Llera Plaza* and back again. *American Criminal Law Review*, *41*, 1189–1276.

Cole, S. A. (2005a). Does "yes" really mean yes? The attempt to close debate on the admissibility of fingerprint testimony. *Jurimetrics*, *the Journal of Law*, *Science*, *and Technology*, *45*, 449–464.

Cole, S. A. (2005b). More than zero: Accounting for error in latent fingerprint identification. *The Journal of Criminal Law & Criminology*, *95*, 985–1078.

Cole, S. A. (2006). Is fingerprint identification valid? Rhetorics of reliability in fingerprint proponent's discourse. *Law & Policy*, *28*, 109–135.

Cole, S. A. (2007). The fingerprint controversy. *Skeptical Inquirer*, *31*(4), 41–46.

Cole, S. A., & Lynch, M. (2006). The social and legal construction of suspects. *Annual Review of Law and Social Science*, *2*, 39–60.

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, *56*, 600–616.

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, *156*, 74–78.

Eng, J. (n.d.). *ROC analysis: Web-based calculator for ROC curves.* Retrieved August 3, 2006, from http://www.jrocfit.org.

Epstein, R. (2002). Fingerprints meet Daubert: The myth of fingerprint "science" is revealed. *Southern California Law Review*, *75*, 605–657.

Faigman, D. L. (2002). Is science different for lawyers? *Science*, *297*, 339–340.

Faigman, D. L., Kaye, D. H., Saks, M. J., & Sanders, J. (Eds.). (2002). *Modern scientific evidence: The law and science of expert testimony* (2nd ed., Vol. 2). St. Paul, MN: West.

Fine, G. A. (2006). *A review of the FBI's handling of the Brandon Mayfield case* (Tech. Rep.). Washington, DC: U.S. Department of Justice Office of the Inspector General.

Galton, F. (1893). *Decipherment of blurred finger prints: Supplementary chapter to "finger prints"*. London: Macmillan.

Haber, L., & Haber, R. N. (2003). Error rates for human fingerprint examiners. In N. K. Ratha & R. M. Bolle (Eds.), *Automatic fingerprint recognition systems* (339–349). New York: Springer-Verlag.

Haber, L., & Haber, R. N. (2006). Letter to the editor. Re: A report of latent print examiner accuracy during training exercises. *Journal of Forensic Identification*, *56*, 493–499.

Heidemann, G. (2006). The principal components of natural images revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 822–826.

Kennedy, D. (2003). Editorial, forensic science: Oxymoron? *Science*, *303*, 1625.

Loftus, E. F., & Cole, S. A. (2004). Contaminated evidence. *Science*, *304*, 959.

Maio, D., Maltoni, D., Cappelli, R., Wayman, J. L., & Jain, A. K. (2002). FVC2000: Fingerprint verification competition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 402–412.

Megreya, A., & Burton, A. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*, 865–876.

Megreya, A., & Burton, A. M. (2007). Hits and false positives in face matching: A familiarity-based dissociation. *Perception and Psychophysics*, *69*, 1175–1184.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). A low-dimensional representation of faces in the higher dimensions of the space. *Journal of the Optical Society of America A*, *10*, 405–411.

Pankanti, S., Prabhakar, S., & Jain, A. K. (2002). On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1010–1025.

Siegel, D. M., Acree, M. A., Bradley, R., Cole, S. A., Faigman, D. L., Fienberg, S. E., et al. (2006). The reliability of latent print individualization: Brief of Amici Curiae submitted on behalf of scientists and scholars by the New England Innocence Project, Commonwealth v. Patterson. *Criminal Law Bulletin*, *42*, 21–51.

Specter, M. (2002, May 27). Do fingerprints lie? *The New Yorker*, 96–105.

Starrs, J. E. (1999). Judicial control over scientific supermen: Fingerprint experts and others who exceed the bounds. *Criminal Law Bulletin*, *35*, 234–276.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*, 71–86.

Vokey, J. R. (2001, June). *LOVE (latent orthogonal visual elements) in pictures.* Paper presented at Lake Ontario Visionary Establishment annual conference, Niagara Falls, Ontario, Canada.

Vokey, J. R., & Higham, P. A. (2004). Opposition logic and neural network models in artificial grammar learning. *Consciousness and Cognition*, *3*, 565–578.

Vokey, J. R., Rendall, D., Tangen, J. M., Parr, L. A., & de Wall, F. B. M. (2004). Visual kin recognition and family resemblance in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *118*, 194–199.

Vokey, J. R., & Tangen, J. M. (2001, June). *Learning an artist's style: What does a pigeon see in a Picasso?* Paper presented at Annual Meeting of the Canadian Society for Brain, Behaviour & Cognitive Science, Université Laval, Québec, Canada.

Wertheim, K., Langenburg, G., & Moenssens, A. (2006a). Authors' response to letter: Letter to the editor. Re: A report of latent print examiner accuracy during training exercises. *Journal of Forensic Identification*, *56*, 500–510.

Wertheim, K., Langenburg, G., & Moenssens, A. (2006b). A report of latent print examiner accuracy during comparison training exercises. *Journal of Forensic Identification*, *56*, 55–93.

Wickens, T. D. (2001). *Elementary signal detection theory*. New York: Oxford University Press.

Wilson, C., Hicklin, R. A., Korves, H., Ulery, B., Zoepfl, M., Bone, M., et al. (2004). *Fingerprint vendor technology evaluation 2003: Summary of results and analysis report* (Tech. Rep. No. NISTIR 7123). Washington, DC: National Institute of Standards and Technology.

Zabell, S. L. (2005). Fingerprint evidence. *Journal of Law and Policy*, *13*, 143–179.