Humans matching fingerprints: Sequence and Size

Matthew B. Thompson[1,2], Jason M. Tangen[1,2], Renee Treloar[1] & Kathleen Ivison[1]

[1]School of Psychology, The University of Queensland
[2]National ICT Australia

Television shows like ‘CSI’ can give the impression that matching crime-scene fingerprints is fully automated. But it is actually humans (fingerprint experts) who ultimately decide whether a crime-scene print belongs to a suspect or not. Despite this fact, there have been no published, peer-reviewed studies directly examining the extent to which experts can correctly match fingerprints to one another. In two experiments presented here we aim to determine the factors affecting accuracy using non-expert participants and test (1) whether the advantage found for the sequential presentation of faces applies to prints and (2) whether the amount of information in a print matters.

INTRODUCTION

The television show ‘CSI’ gives viewers the impression that matching crime-scene fingerprints is fully automated. Detectives simply enter an image of a crime-scene fingerprint into a computer and sophisticated pattern recognition software instantly matches the print to a person on file. In reality, it is actually humans (fingerprint experts) who ultimately decide whether a crime-scene print belongs to a suspect or not (Vokey, Tangen, & Cole, 2008). Despite this fact, there have been no published, peer-reviewed studies directly examining the extent to which experts can correctly match fingerprints to one another (Cole, 2005, 2007; Haber & Haber, 2003; Loftus & Cole, 2004)

Professional fingerprint experts (FPEs) aim to attribute a crime-scene or “latent” print to a “fully-rolled” ten-print set taken from a suspect or a database. A recent National Academy of Sciences (NAS; 2009) report concluded that most of the forensic disciplines rely on subjective assessment of matching characteristics and that research on the limits and measures of forensic expert’s performance is sorely needed. The NAS report also notes that the forensic domain could benefit from research conducted in other areas such as evaluation of observer performance in diagnostic medicine and from the findings of cognitive psychology on the potential for bias and error in human observers.

Before presenting our research we discuss previous investigations involving fingerprint experts. Busey and Vanderkolk (2005) compared expert and novice fingerprint examiners in an EEG study. They examined whether the face inversion effect (i.e. when presenting a face upside-down it dramatically affects its recognition and increases the delay of the N170 brain response) found for objects of expertise would also be shown for people viewing upside-down fingerprints. They found an inversion effect for expert fingerprint examiners but not for novices, suggesting that experts perceive prints configurally (the patterns as a whole) rather than featurally (the tiny features or minutuae).

Context appears to have a dramatic effect on the accuracy of FPEs. Dror, Charlton and Peron (2005) used a highly-publicized case of exposed fingerprint error to determine whether biasing information could lead an examiner to change prior judgments. They covertly evaluated five FPEs, with an average of 17 years of experience, who consented to being tested at an unknown time over twelve months. They obtained a set of prints that each FPE had declared previously as a match within the previous five years. The five FPEs were each given a print to identify by a colleague, who advised them that the fingerprints were from a famous case of misidentification by the FBI for the 2004 Madrid train bombings. One reported that the prints matched, three reported that the prints did not match, and one reported inconclusive.

Unbeknownst to the FPEs, however, the prints they were asked to identify were taken from their own previous case history where they had previously declared them a match. With four of the five examiners subsequently changing their previous judgment of the prints as matching, it is clear that top-down, contextual influences can affect the judgments of FPEs.

Although there is no good research on the matching accuracy of fingerprint experts, research has been conducted on novices. Vokey, Tangen and Cole (2009) had naive undergraduate participants match fingerprints on a computer screen. Their fingerprint stimuli were obtained from university volunteers who were asked to ink each of their fingers (with a professional compound) and leave two prints of each finger on paper. Two fingerprints were presented to participants side-by-side and responded either “match” or “non-match.”

Through signal detection analysis Vokey et al. (2009) showed that naive undergraduate participants can match fingerprints above chance and that their matching accuracy can vary as a function of source finger type (thumb, index, middle, etc.). Vokey et al. were the first to determine a baseline measure of difficulty for the task of matching fingerprints and first to assess some of the factors that may influence that difficulty. Their stimuli set, however, is different from that used by fingerprint experts who match latent crime-scene prints to fully-rolled ten-print cards.

In a follow-up study Tangen and Vokey (2010) replicated the Vokey et al. (2009) study using actual forensic latent prints, acquired from Queensland Police Service training materials, with their corresponding matching prints from full-rolled ten-print cards. They found that, although the overall level of accuracy decreased slightly relative to the Vokey et al. results, participants were still able to reliably match latent crime-scene prints to full-rolled ten-print cards.

The above two studies provide a baseline measure of novice matching accuracy and can be eventually used to compare against expert performance. The next step in this program of research is to investigate the factors that affect people's matching accuracy. In two experiments presented here we use novice participants and test (1) whether the advantage found for the sequential presentation of faces also applies to fingerprints and (2) whether the amount of information in a fingerprint matters.

EXPERIMENT 1

Research on the accuracy of eyewitness identification shows that the number of false alarms is reduced when face stimuli are presented sequentially (one at a time) rather than simultaneously (all at once; Steblay, Dysart, Fulero, & Lindsay, 2001). We aim to replicate this effect using a different face database and also determine whether the advantage seen for sequential presentation of faces holds when fingerprints are presented sequentially rather than simultaneously.

Method

Ninety-six psychology undergraduates from The University of Queensland participated in exchange for course credit. A 2 × 2 × 2 within-groups factorial design was employed: stimulus type (fingerprint or face), presentation mode (simultaneous or sequential), and trial type (target or distractor).

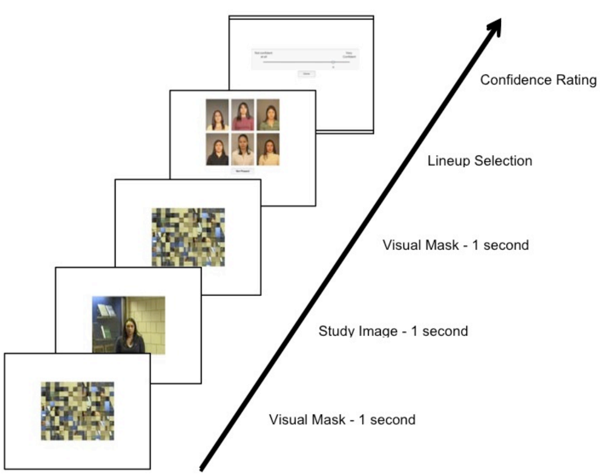

Figure 1. Example of the simultaneous presentation procedure with faces.

Participants asked to judge whether prints matched or not using a confidence rating scale ranging from 1 (sure different) to 12 (sure same). In the simultaneous condition participants viewed a study image for one second followed by six images where the target image may or may not be present (see Figure 1). In the sequential condition participants viewed a study image for one second followed by six individual lineup images where the target may or may not be present (see Figure 2). If the participant declared a match on any of the lineup images the remaining images would not be displayed and the next trial would commence.

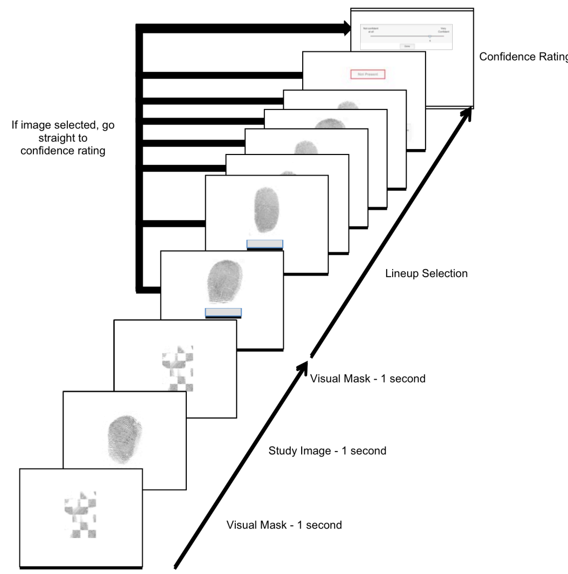

Figure 2. Example of the sequential lineup procedure with prints.

The face stimuli included an equal number of male and female Caucasian faces and the finger stimuli included an equal number of each finger type (thumb, index, middle, ring, little) across all trial types and presentation modes. We measured the proportion of hits (i.e., declaring a "match" when the prints, in fact, match) and false alarms (i.e., declaring a "match" when the prints do not, in fact, match).

Fingerprint stimuli were taken from the database of prints used in Vokey et. al. (2009). The set consisted of full ten-print sets obtained from 125 individuals on two separate occasions (total of 1250 fingerprints per occasion). The faces were obtained from the Face Recognition Grand Challenge Database (2006). In total, we used images of 260 people (130 female and 130 male).

Participants completed 80 trials (four blocks of 20 trials), where half were fingerprint lineups and half were face lineups. Within each of these fingerprint and face lineups, 20 were presented simultaneously and 20 were presented sequentially. On each trial, a visual mask was presented on the screen for 1 second prior to and following the presentation of the study image. The masks consisted of 300 scrambled blocks of the original image and the fingerprint masks consisted of 96 scrambled blocks of the original image (Busey & Vanderkolk, 2005). The study image (face or fingerprint) was also presented for 1 second and participants were asked to remember it and were then presented with a lineup and were instructed to select the face or fingerprint that they were asked to remember, or indicate “Not Present” if the face or print was not in the lineup.

Within each set of 20 trials, 10 of the study images had their corresponding image present in the subsequent lineup and 10 were not present. The order of the four sets of 20 trials was counterbalanced across participants, and prior to each block of 20 trials.

Results

Participants indicating a correct match on target trials were classified as hits, and indicating a match on distractor trials were classified as false alarms.

Hit rates were significantly higher for faces than for prints ($p < .001$) and there were fewer false alarms for faces than for prints ($p < .001$). Hit rates for simultaneous lineups were significantly higher than hit rates for sequential lineups ($p < .001$). False alarm rates for sequential lineups were lower than for simultaneous lineups ($p < .001$).

A' was calculated to give a measure of participant’s ability to discriminate targets from distractors. A' scores close to 1 represent good discriminability, scores around .5 represent chance levels of discrimination, and scores below .5 represent poor discriminability (Pallier, 2002). Participants showed better discrimination of faces than of prints ($p <.001$; supporting H2). There was no difference in discrimination between sequential and simultaneous lineups.

Participants were more conservative for faces than prints and were more conservative for sequential lineups than for simultaneous lineups. Participants were more confident in target trials than distractor trials ($p = .001$), and more confident for simultaneous lineups than sequential lineups ($p < .001$). Participants were just as confident for faces as they were for prints ($p = .21$) despite the finding above that they were much worse at discriminating prints than faces.

EXPERIMENT 2

How the amount of information or the size of a fingerprint affects accuracy is currently unknown. Previous research has indicated that fingerprint examiners often examine latent fingerprints containing as little as 20% of their original information (Haber & Haber, 2003). We explore the role of print size on fingerprint identification performance by manipulating amount of visual information contained in fingerprints. We predict that large prints will elicit more false alarms (i.e. declaring a “match” on distractor trials) than small prints (Vokey et al., 2009)

Method

One hundred and twelve psychology undergraduates from The University of Queensland participated in exchange for course credit. A 2 × 2 × 2 within-groups factorial design was employed: trial type (target or distractor), fingerprint size (small or large) and finger type (thumb, index, middle, ring or little finger). Equal proportions of each finger type across all trial types and fingerprint sizes were used. Stimuli were taken from the database of prints used in Vokey et. al. (2009) consisting of full ten-print sets obtained from 125 individuals on two separate occasions.

2D Gaussian masks were applied to the fingerprint images to manipulate the surface area of the print presented. Each print was cropped to the width of the fingerprint image resulting in a square grayscale matrix (260 × 260 pixels, ranging from 1-255). This matrix was then multiplied by different probability distribution functions to produce the large and small masks resulting in large masked prints that were 220 × 220 pixels and small masked prints that were 110 × 110 pixels.

Two fingerprints were presented side-by-side on the computer screen. One was a masked print (either large or small), and the other was a full (unmasked) print. Participants completed a total of 80 trials where they compared pairs of prints. For half of the trials, the two fingerprints presented were a match (targets) and in the other half of trials the prints did not match (distractors).

Participants were asked to judge whether the prints matched or not using a confidence rating scale (see Figure 3). The scale ranged from 1 (sure different) to 12 (sure same).

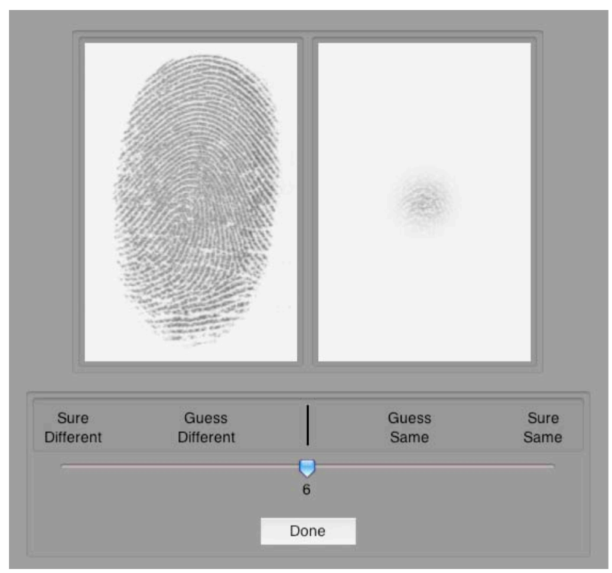

Figure 3. Screenshot of experiment showing the target print on the left and a print from the small condition on the right. A slider for matching and for confidence is below.

Results

The analysis revealed significant main effects of both size ($p < .001$), and finger type, ($p < .001$). Index fingers were better discriminated than all other fingers and little fingers were discriminated less well than all other fingers. The hit rate was higher and the false alarm rate was lower for large prints compared to small prints, and the proportion of hits and false alarms varied as a function of finger type and area ($p < .001$).

DISCUSSION

We set out to determine the factors affecting the matching accuracy of non-expert participants and test (1) whether the advantage found for the sequential presentation of faces applies to prints and (2) whether the amount of information in a print matters.

Experiment 1 found that novices were not able to identify prints from a line-up but were able to identify faces from a line-up. Surprisingly, participants were just as confident in their judgments of faces as they were for prints despite the finding that they were much worse at identifying prints than faces. Participants were significantly more liberal in their judgments for prints than faces and so were more inclined to select a print out of a lineup than select the "Not Present" option. They were able to discriminate faces reasonably well and they were significantly more conservative in their judgments for faces than prints.

Contrary to our expectations, there was no significant difference between the two presentation modes for faces or prints. That is, whether identifying faces or fingerprints, there was no difference in accuracy between simultaneous and sequential line-ups. This may be due to the differences between the face stimuli we used compared to that used in previous research.

The marked differences between the study images and the targets for our face stimuli (often taken weeks apart and in different locations and orientations) appears to make discrimination difficult and may have created a floor effect, leading to a failure to replicate the benefit for sequential presentation. Although a meta-analysis has shown that the effect is reduced in when studies approximate for real-world conditions such as those present in our face stimuli (Steblay, et al., 2001). Failure to find a benefit for sequential presentation of prints may be due to participant's floor levels of fingerprint identification accuracy.

Experiment 2 showed that novices discriminate large prints significantly better than small prints. It appears that, at least in this case, novices are more accurate when prints have a larger surface area. But because we only had two discrete size conditions it is unclear whether this effect is linear. It may be that prints in the large condition were closer in size to smaller prints from previous research.

Most interestingly, accuracy also varied as a function of finger type. This is surprising given that such a small about of the original print was used in this experiment, even in the large condition (see Figure 3). Also, when inspecting a person's prints from the five finger types side-by-side, there appears to be little difference in the amount of information between them. There is currently no good theory to explain this result. It may be that fingers used more often may acquire more distinguishing characteristics. Further research is required but it is clear that manipulating the size of a print is just one way to vary the amount of information present and further studies could examine different manipulations of information content.

CONCLUSION

It appears that novices find it difficult to remember fingerprints and identify them from a line-up. This may be the reason for finding no difference between simultaneous and sequential line-ups in the present study. It is also clear that novices match large prints more accurately than small prints and that, despite finger size being controlled for, some finger types are discriminated more accurately than others. Further research should investigate whether fingerprint experts perform differently than novices in both these experiments.

Psychological and human factors investigations into the performance and support of expert and novice fingerprint examiners is sparse. Much of what is already known about human error, vigilance, and decision making from domains such as healthcare and aviation can be applied to the forensic domain to improve the accuracy of fingerprint, shoeprint, face, voice and DNA matching.

ACKNOWLEDGEMENTS

This research was supported by strategic funding from the School of Psychology at The University of Queensland.

REFERENCES

Busey, T., & Vanderkolk, J. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45*, 431-448.

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

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

Face Recognition Grand Challenge Database (2006). Retrieved 1 September 2009, from www.frvt.org/FRGC

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.

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

National Academy of Sciences (2009). *Strengthening Forensic Science in the United States: A Path Forwar.* Washington, DC: The National Academies Press.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous linup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459-473.

Tangen, J. M., & Vokey, J. R. (submitted). Fingerprint identification: On the psychophysics of matching latents to ten-print patents. *Canadian Journal of Experimental Psychology.*

Vokey, J. R., Tangen, J. M., & Cole, S. A. (2008). On the preliminary psychophysics of fingerprint identification. *The Quarterly Journal of Experimental Psychology, 62*, 1023-1040.